Neural correlates of economic decision-making in the regions and projections of the frontal cortex



Caitlin Ashcroft

Oriel College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2024

Life is a sum of all your choices — Albert Camus

Abstract

Economic decision-making relies on the ability to evaluate and integrate information in our surroundings to guide goal-directed actions. Neurons across the frontal cortex have been found to heterogeneously encode a wide range of sensory, behavioural, and cognitive variables relevant to decision-making and reward-motivated behaviour. However, precisely how the encoding of economic variables is structured across frontal regions and, more broadly, in the network of brain regions that support decision-making remains unclear. This thesis explores the structure of the neural code underlying economic decision-making by studying the neural correlates of decision-making across multiple regions and projections of the frontal cortex. We established a novel economic decision-making task in mice and combined this with large-scale electrophysiological recordings to monitor the activity of thousands of neurons across the frontal cortex during decision-making behaviour. Furthermore, we implemented an in vivo optotagging protocol to isolate neurons projecting to the dorsomedial striatum and claustrum - two major hubs of frontal inputs - to investigate how the coding of economic variables maps onto specific frontal projection pathways. These data reveal a novel spatial structure underlying the encoding of fundamental economic variables in the frontal cortex and indicate that decision-relevant information is selectively routed to specific downstream targets. Together, the results of this thesis suggest that the encoding of economic variables spans multiple scales of representation in the frontal cortex.

Preface

This thesis describes my work investigating the neural correlates of economic decision-making across the regions and projections of the frontal cortex.

All the work described in this thesis was performed in the laboratory of my supervisor, Dr. Armin Lak, in the Department of Physiology, Anatomy, and Genetics at the University of Oxford.

In Chapter 1, I provide an overview of relevant literature and motivate the research questions this thesis seeks to address.

In Chapter 2, I outline the key methods and materials used in this study.

In Chapter 3, I describe the novel visually-guided decision-making task used to study economic decision-making behaviour in this project. This chapter outlines the key behavioural results indicating that this paradigm is appropriate for studying value-based decision-making behaviour in mice.

In Chapter 4, I describe the results obtained from large-scale electrophysiological recordings across the frontal cortex of mice performing this task. These data revealed distinct spatial gradients for the encoding of key decision variables during economic decision-making behaviour.

In Chapter 5, I describe the results obtained from combining large-scale electrophysiological recordings and in vivo optotagging to monitor the activity of projection-defined neurons during decision-making behaviour. These data demonstrate that projection-defined populations of frontal neurons, specifically those projecting to the dorsomedial striatum and claustrum, differentially encode decision variables, beyond the signals observed at a population level.

In Chapter 6, I summarise the key results from the previous chapters and consider the wider implications and possible future directions arising from this work.

Summary of contributions

I, Caitlin Ashcroft, confirm that the work presented in this thesis is my own. Any work contributed by collaborators or information derived from other sources has been appropriately indicated throughout. While all writing in this thesis is my own, this project relies on significant contributions from others. Throughout the text, the pronoun 'we' will be used, indicating myself and the team below.

Antara Majumdar: Contributed behavioural data from 15 mice (Chapter 3) and large-scale electrophysiological data from 12 mice (Chapter 4). Antara also assisted with data analysis, particularly behavioural analysis (Chapter 3)

Matthias Fritsche: Assisted with data analysis, particularly the development of the linear decoding pipeline (Chapter 4).

Lukasz Bijoch: Assisted with data analysis, particularly waveform correlation (Chapter 5)

Lauren Strickland: Contributed behavioural data from 3 mice (Chapter 3)

Niamh Walker: Performed the histological processing, imaging, and data analysis for the anatomical tracing experiment (Appendix A.4), under my supervision

Contents

1 Introduction **14**

1.1 The role of frontal cortex in higher-order cognition and decision-making 15

1.2 Clinical implications of disordered decision-making 16

1.3 Functional specialisation across the sub-regions of the frontal cortex during decision-making behaviour 17

1.4 Single neurons in the frontal cortex encode decision-relevant information 19

1.5 Neural encoding of economic variables 20

1.6 Widespread and heterogenous representation of decision variables across frontal cortical regions 22

1.7 Projection-defined signals in frontal cortex 24

1.8 Frontal inputs to striatum and claustrum 26

1.9 Thesis overview 28

2 General methods and materials **29**

2.1 Subjects 29

2.2 Stereotaxic surgery 31

2.2.1 General surgical preparation and post-operative care 31

2.2.2 Head-plate surgery 31

2.2.3 Intracranial injections and optic fiber implantations 32

2.2.4 Tracing surgeries 33

2.3 Behaviour 33

2.3.1 Materials and apparatus 33

2.3.2 Visual economic decision-making task 33

2.3.3 Behavioural training 34

2.3.4 Behavioural inclusion criteria . . . 35
2.4 Large-scale electrophysiological recordings . . . 35
2.4.1 Electrophysiological recordings . . . 35
2.4.2 Electrophysiological data processing . . . 36
2.5 Optotagging projection-defined neurons . . . 37
2.5.1 Optotagging protocol . . . 37
2.5.2 Optotagged neuronal selection criteria and validation . . . 38
2.6 Histology and probe alignment . . . 39
2.6.1 Histological processing . . . 39
2.6.2 Probe tract alignment . . . 40
2.6.3 Anatomical tracing image analysis . . . 40
2.7 Data analysis . . . 41
2.7.1 Population decoding . . . 41
2.7.2 Single cell analysis . . . 45

3 Establishing a economic decision-making task in mice 46
3.1 Introduction . . . 46
3.1.1 Task outline . . . 48
3.2 Results . . . 50
3.2.1 Task performance reflects differences in expected value . . . 50
3.2.2 Response times reflect differences in expected value . . . 52
3.2.3 Accounting for the effects of visual or motor biases on task performance . . . 54
3.2.4 Task acquisition . . . 56
3.3 Discussion . . . 58
3.3.1 Limitations and future directions . . . 59
3.4 Summary . . . 61

4 Representation of decision-variables across frontal cortical regions during decision-making behaviour 62
4.1 Introduction . . . 62
4.2 Results . . . 63

4.2.1 Distribution of recorded neurons across frontal cortex 63
4.2.2 Individual neurons are differentially modulated by task conditions at stimulus onset 66
4.2.3 Establishing the appropriate time window for investigating stimulus responses 66
4.2.4 Population activity across the frontal cortex reflects differences in expected value 69
4.2.5 Expected value and risk encoding in frontal cortical regions . . 69
4.2.6 Spatial distribution of variable encoding in frontal cortex . . . 72
4.2.7 Lateralised neural responses 74
4.2.8 Interaction of stimulus location and economic variable encoding 76
4.3 Discussion 79
4.3.1 Limitations and future directions 82
4.3.2 Summary 85

5 Representation of decision-variables in projection-defined frontal cortical neurons during decision-making behaviour 86
5.1 Introduction 86
5.2 Results 88
5.2.1 In vivo optotagging identified populations of DMS and CLA-projecting frontal neurons 88
5.2.2 Distribution of projection-defined neurons across frontal cortex 91
5.2.3 DMS- and CLA-projecting frontal neurons differentially encode economic variables 92
5.2.4 Differential encoding of EV and Risk in DMS- and CLA-projecting neurons in dorsomedial frontal cortex 95
5.3 Discussion 97
5.4 Limitations and future directions 100
5.5 Summary 101

6 Discussion 103
6.1 Economic decision-making in mice 104

6.2 Spatial gradients in neural encoding of economic variables 105
6.3 Projection-specific coding 107
6.4 Future directions 108
6.5 Conclusion 110

A Appendix A **111**
A.1 Two-stimulus trials 111
A.2 Identifying task-responsive neurons 113
A.3 Correlating variable encoding 114
A.4 Anatomical characterisation of dual-projecting neurons 116

B Appendix B - Developing an optotagging set-up to target projection-defined frontal cortical neurons in vivo **119**
B.1 Introduction 119
B.2 Optimising surgical procedures 122
B.3 Testing the experimental set-up 124
B.4 Combining optotagging and large-scale electrophysiological recordings 126
B.5 Waveform analysis to validate laser-evoked neural activity 128
B.6 Establishing selection criteria to identify laser-responsive neurons . . 130
B.7 Collision testing 133
B.8 Discussion 136
B.8.1 Limitations 137
B.8.2 Summary 139

Bibliography **142**

List of Figures

2.1 Manual curation of spikesorted data 37
2.2 In vivo optotagging criteria for identifying CLA- and DMS-projecting frontal cortical neurons 39

3.1 A visually-guided decision-making task to probe economic decision-making in mice 50
3.2 Choice accuracy scales with expected value (EV) 51
3.3 Performance on two stimulus trials 54
3.4 Controlling for visual and motor biases 55
3.5 Task acquisition 57

4.1 Large scale recordings across frontal cortical regions identified individual neurons modulated by task variables 65
4.2 Population activity at stimulus onset reflects differences in expected value 68
4.3 Binary decoding of expected value and risk across frontal cortical regions 71
4.4 Distinct spatial gradients for EV and risk coding across frontal cortex 73
4.5 Encoding of stimulus location in frontal cortex 75
4.6 Ipsilateral and contralateral coding of EV and risk 78

5.1 In vivo optotagging identified populations of DMS and CLA-projecting neurons in the frontal cortex (FCx) 89
5.2 Distribution and encoding of DMS- and CLA-projecting frontal cortical neurons 93
5.3 Encoding in dorsomedial frontal cortex neurons projecting to DMS and CLA 96

A.1 Performance on two stimulus trials 112
A.2 Identifying task-driven responses across task epochs 113
A.3 Stimulus location, EV, and risk coding correlation 115
A.4 Anatomical tracing confirmed the existence of divergent frontal projections to DMS and CLA 117

B.1 Refinement of surgical strategy enabled reliable targeting of opsin expression and optic fiber implantation above CLA and DMS 124
B.2 Testing the experimental set up for in optotagging 125
B.3 Laser delivery drove a robust change in firing rate in a subset of recorded neurons 127
B.4 Example waveform correlation between spontaneous and laser-evoked spikes 129
B.5 In vivo optotagging criteria for identifying CLA- and DMS-projecting frontal cortical neurons 133
B.6 Collision testing to validate antidromic spikes 136

List of Tables

1.1 Lesion and neuroimaging evidence supporting functional specialisation across the sub-regions of frontal cortex in the context of decision-making behaviour 18

2.1 Subject overview 30

3.1 Accuracy differences across trial types 52
3.2 Response times across stimulus conditions 53
3.3 Response time reflects differences in expected value 53

4.1 Distribution of recorded and task-modulated neurons across frontal regions 64

5.1 Distribution of DMS-projecting neurons across frontal regions 91
5.2 Distribution of CLA-projecting neurons across frontal regions 92

A.1 Mean accuracy across two stimuli trial conditions (different EV) . . . 112
A.2 Mean accuracy across two stimuli trial conditions (samet EV, different risk) 113

B.1 Projection-specific optotagging criteria in existing literature 130

Chapter 1

Introduction

Making decisions is a fundamental part of life. From selecting what to eat on a restaurant menu to deciding how to invest money, we are constantly faced with choices between various options, often with highly variable and uncertain outcomes. To make 'good' choices in such conditions, individuals must consider the potential value of different options, and the probability that each will result in a desirable outcome, to select actions that align with their goals. The concepts of 'expected value' and outcome uncertainty, or 'risk', have been the bedrock of economic and psychological theories of decision-making behaviour for decades[1–4]. A large body of work from economists, psychologists, and neuroscientists has sought to shed light on how these variables shape decision-making behaviour in a variety of real-world and experimental contexts.

The ability of humans, non-human primates, and other mammals to make economic decisions relies, at least in part, on their elaborate and highly interconnected frontal cortex. There is ample evidence that the frontal cortex plays a critical role in decision-making and goal-directed behaviour, from the systematic disruptions caused by frontal cortical lesions[5–11] to the observation that single neurons across the frontal cortex encode a wide array of sensory, cognitive, and behavioural variables relevant to decision-making[12–17]. However, precisely how decision-relevant information is represented across the sub-regions of the frontal cortex and, more broadly, within the network of brain regions that support decision-making remains unclear. Therefore,

the key question this thesis seeks to address is:

How are the neural correlates of economic decision-making structured across the regions and projections of the frontal cortex?

1.1 The role of frontal cortex in higher-order cognition and decision-making

It might surprise a modern-day neuroscientist to learn that the frontal cortex was once considered broadly obsolete. Early observations that large portions could be damaged or ablated without causing severe mental impairments or behavioural changes[18–20] were interpreted as evidence that the frontal cortex was essentially 'cognitively silent'. This early view of minimal frontal cortex function likely contributed to the widespread use of invasive surgery to treat psychiatric disorders in the late 19th and early 20th centuries[21]. However, these early observations starkly contrast with other reports of cognitive changes associated with frontal cortical damage, perhaps most famously in the case of Phineas Gage over 150 years ago. In a turn of events that would make any Occupational Health department shudder, a projectile tamping rod destroyed a significant portion of Gage's frontal lobe, turning him from a diligent and mild-mannered railroad worker to an impulsive, emotionally volatile, and disinhibited psychiatric case study. His physician would later refer to this array of symptoms as 'frontal lobe syndrome'[5].

It is now widely recognised that while patients with frontal lesions may perform normally on tests of intelligence, motor control, sensation, and perception, such individuals often display profound deficits in other domains such as planning[6;22;23], problem-solving[7;9], organisation[8;24;25], and social and emotional regulation[26]. Together, these functional disruptions point to a central role of the frontal cortex in higher-order cognition and executive control.

Patients with damage to their frontal cortex also display notable deficits in their

ability to make decisions and carry out goal-directed actions. For example, Bechara et al.(1994)[10] trained human participants in a risky decision-making task - the Iowa Gambling Task - in which participants were required to make a series of card selections resulting in the gain or loss of monetary rewards. While healthy subjects quickly developed a preference for the low-risk option that guaranteed an overall profit, patients with frontal lesions - specifically in their ventromedial prefrontal cortex - developed a consistent and sustained preference for riskier options that resulted in an overall loss. This study, along with other work from the groups of Antonio Damasio and Trevor Robbins[11;27–31], suggests that damage to the frontal cortex can hinder the ability to evaluate the utility of different options and outcomes and to update expectations and choices accordingly. Thus, the frontal cortex appears to occupy a crucial position in the neural computations that underpin decision-making.

1.2 Clinical implications of disordered decision-making

Biased or disordered decision-making is a feature of several neurological and psychiatric disorders, including addiction[32;33], obsessive-compulsive disorder (OCD)[34], schizophrenia[35], and depression[36]. Evidence suggests that the deficits in decision-making observed in these conditions may arise from aberrant representations of certain features of decision-making processes. For example, in the case of drug addiction, previous research has identified excessive attentional bias towards drug-related cues[37;38] and underestimation of, or insensitivity to, negative outcomes[39;40]. This abnormal salience attribution and impaired response inhibition is thought to underlie the development of cravings, compulsive and excessive drug use, and impaired self-awareness that characterise substance use disorders[41;42]. Similarly, major depression is associated with heightened salience of negative outcomes and diminished responses to positive outcomes[43;44]. These findings suggest that developing a better understanding of how neural computations underpinning decision-making are altered in psychiatric disorders could improve our ability to support individuals managing these conditions and potentially identify novel treatment targets. However, such advances

require a better understanding of the mechanisms underpinning decision-making in healthy individuals, enabling us to assess how these mechanisms are disrupted in psychiatric disorders.

1.3 Functional specialisation across the sub-regions of the frontal cortex during decision-making behaviour

The frontal cortex is not a single homogeneous structure. Rather, it is comprised of several sub-regions with distinct cytoarchitecture and input-output connectivity[45–49]. Focal lesion studies, supported by neuroimaging literature, have indicated that different frontal cortex regions may play distinct roles in decision-making behaviour. For example, Buckley et al (2009)[50] trained macaques in a decision-making task analogous to the Wisconsin Card Sorting Task[7] and tested the relative effect of lesions in different frontal regions including the dorsolateral prefrontal cortex (DLPFC), anterior cingulate cortex (ACC), or orbitofrontal cortex (OFC). The authors observed that damage to the DLPFC appeared to specifically impair the maintenance of task rules in working memory, while OFC lesions disrupted animals' ability to update the value of rules based on experienced rewards. Lesions to the ACC, on the other hand, were found to selectively impair response selection and the evaluation of choice-outcome associations. These findings suggest that different sub-regions of the frontal cortex contribute to distinct features of decision-making in this task.

A large body of lesion and neuroimaging evidence points towards functional specialisation across the structures that comprise the frontal cortex in the context of decision-making behaviour (Table 1.1). In general terms, the dorsolateral PFC has been largely implicated in the maintenance and manipulation of task-relevant information in working memory[23;25;51;52], as well as for rule-learning[7;9] and planning[11]. The ventromedial prefrontal cortex (vmPFC), and the adjacent orbitofrontal cortex (OFC), have been highlighted as key structures for value-based decision-making. Neuroimaging studies have frequently identified value signals in the vmPFC that broadly

scale with the expected value or utility of choices[53–57]. Glascher et al (2009)[57], for example, used fMRI to image the frontal cortices of human subjects performing a simple reward-related decision task and observed activity in vmPFC that tracked expected future rewards in both stimulus- and action-driven task variants. Consistent with this, patients with vmPFC damage often display impairments in their ability to make value-maximising choices[58;59] or establish consistent preferences between options[60]. The OFC, the medial portions of which are sometimes grouped with the vmPFC[61], has also been implicated in value-based decision-making, particularly for learning and updating stimulus-value associations and reversal learning[10;11;50;62;63]. In the case of the anterior cingulate cortex (ACC), lesion and neuroimaging approaches have pointed towards a role in error detection[64], conflict monitoring[65], effort-based decision-making[66–68] and the learning the value of actions[69;70].

Brain Region	Studies	Function
DLPFC	Milner et al., 1963; Petrides and Milner, 1982; Shallice, 1982; Owen et al., 1990; Shallice and Burgess, 1991; Chao and Knight, 1997; Manes et al.2002; Buckley et al., 2009; Tsuchida and Fellow, 2009	Working memory Rule learning Planning Attention
VMPFC/OFC	Behcara et al., 1994; Bechara et al, 1997; Bechara et al, 2000; Knutson et al, 2005; Kable and Glimcher, 2007; Hare et al, 2008; Glasher et al, 2009; Buckley et al, 2009; Walton et al., 2010; Levy and Glimcher, 2011	Expected and subjective value Learning/updating stimulus-value contingencies Stable preference Cognitive flexibility Reversal learning
ACC	Walton, Bannerman and Rushworth, 2002; Walton et al., 2003; Roger et al., 2004; Kennerley et al, 2006; Marsh et al., 2007; Cohen et al., 2008; Behrens et al, 2008; Buckley et al., 2009	Monitoring gains/losses Error detection Conflict Effort-based decision-making Learning action value

Table 1.1: Lesion and neuroimaging evidence supporting functional specialisation across the sub-regions of frontal cortex in the context of decision-making behaviour

Overall, region-specific frontal lesions and the observation that activity in these areas

correlates with distinct features of decision-making support functional specialisation across the structures that comprise the frontal cortex. However, it is important to note that such approaches only offer a coarse means of investigating the function of specific neural structures due to the restricted spatial and temporal resolution of non-invasive imaging methods and the often diffuse nature of lesions, particularly in humans. Consequently, these approaches cannot provide insight into the cellular mechanisms underpinning specific decision processes or pinpoint the circuits in which they occur. Thus, determining the contribution of different frontal regions to distinct features of decision-making and goal-directed behaviour necessitates experimental approaches that afford greater specificity and can be used to study neural processes at the level of single cells. It is in part because of this, alongside advances in the tools available for monitoring and manipulating neural activity in vivo, that the field has moved towards the use of electrophysiological and optical approaches to provide a more direct means of investigating the neural basis of decision-making behaviour.

1.4 Single neurons in the frontal cortex encode decision-relevant information

In their influential review, Miller and Cohen (2001)[71] suggest that the role of the frontal cortex in executive control and higher-order cognition stems from "the active maintenance of patterns of activity in the PFC that represent goals and the means to achieve them". In other words, neural activity in the frontal cortex encodes key information necessary for guiding actions to achieve specific goals. This view of frontal cortical function is supported by a large body of work studying single neuron dynamics in the frontal cortex of non-human primates and rodents during decision-making behaviour.

Early evidence that single neurons in the frontal cortex encode cognitive variables came from the observation that neurons in the dorsolateral PFC appeared to represent working memory through persistent firing during the delay period of a spatial task[72–74]. This indicated that sustained activity in frontal cortical neurons can act

as a bridge between stimulus and action by maintaining relevant information 'online' until it can be used to drive an appropriate response. At a similar time, frontal neurons in several sub-regions (LPFC, ACC, OFC) were identified that displayed a range of reward-associated responses, reflecting reward delivery or omission, reward magnitude, predicted reward, and reward preference[13;75–79]. For example, Tremblay and Schultz (1999)[13] identified OFC neurons in macaques that responded to reward-predicting signals during a spatial delayed-response task and were differentially modulated by the relative value of the rewards available. This suggests that the activity of individual neurons in the frontal cortex tracks expected and received rewards.

Another critical insight into the coding properties of frontal neurons came from the demonstration that single neurons in the frontal cortex can encode abstract rules[80]. The authors trained monkeys on a task that required them to switch flexibly between two rules: a 'match' rule when two successively presented visual cues were the same and a 'non-match' rule when the cues were different. Neural recordings in DLPFC, VLPFC, and OFC revealed strong coding of these abstract rules, with the activity of most neurons displaying activity that reflected the current rule. This result demonstrated that frontal neurons encode abstract cognitive variables that support flexible goal-directed behaviour. Thus, neurophysiological approaches examining the response profiles of individual neurons in the frontal cortex suggest that frontal neuronal activity signals a range of sensory and cognitive variables relevant to decision-making.

1.5 Neural encoding of economic variables

The orbitofrontal cortex has been highlighted as a critical structure for economic decision-making. Neurons within OFC have been found to signal a range of economic variables such as expected value[12;81], chosen value[12;82], reward risk[83], preference[13;14], and decision confidence[15]. In a seminal study, Assad and Padoa-Schioppa (2006)[12] recorded neurons in the orbitofrontal cortex of monkeys performing an economic decision-making task. In this task, monkeys were trained to choose between two types of juice offered in differing amounts. Three groups of OFC neurons were

identified whose activity reflected offered value, chosen value, and chosen juice type: all essential variables for the computations underpinning optimal economic choices in this task. These responses were independent of sensorimotor contingencies, suggesting that these neurons encoded abstract economic variables rather than stimuli and/or actions[84]. These observations indicate that OFC is a critical structure for the abstract computations that underpin economic choices.

Despite this emphasis on the orbitofrontal cortex as a key structure for economic decision-making, encoding of economic variables has also been observed in other frontal cortical sub-regions in non-human primates and rodents. For instance, Kennerley et al. (2009)[85] sought to investigate how neural encoding of decision variables compares across different structures in the frontal cortex by simultaneously recording the activity of individual neurons in ACC and LPFC, as well as OFC, in macaques performing a decision-making task. They identified neurons encoding reward probability, magnitude, and cost across all three regions studied. However, they observed a particular over-representation and multiplexing of these variables in ACC, suggesting that the ACC plays a key role in integrating decision variables to guide action selection. This is consistent with observations that neurons in ACC represent a range of action- and reward-related information in both primates[86–91] and rodents[82;92]. For instance, Akam et al. (2021)[92] used calcium imaging to monitor the activity of ACC neurons in mice performing a sequential decision-making task and observed a rich representation of decision variables, including current state, action-state predictions, and surprise signals that disrupted future state transitions when optogenetically manipulated. Together, this work suggests that neural representations in ACC may provide a comprehensive overview of state space to generate and update reward expectations and initiate actions to support optimal choice behaviour. However, understanding the role of the ACC in decision-making is complicated by the observation that the activity of ACC neurons often reflects a wide array of variables depending on the specific tasks used across different studies, earning it the tongue-in-cheek nickname 'the Neuroscientist's Rorschach Test'[93].

In addition, recent work examining the neural correlates of decision-making in rodents

has also implicated the prelimbic cortex (PL), particularly for decision-making under risk and updating value representations based on changing reward contingencies. For example, Passecker et al (2019)[16] trained rats in a two-armed bandit gambling task and observed that the activity of a subset of PL neurons predicted upcoming choices during outcome evaluation. Inhibiting these neurons was observed to promote risky decisions, even during highly unfavourable reward contingencies. This suggests that PL may support optimal choices by monitoring outcomes and updating the predicted value of available options[17]. Other studies have also reported value predictive signals in PL neurons[94], and observed that lesion or inactivation of PL renders animals insensitive to the value of rewards[95] and promotes risky choices[96].

In summary, single neurons across different frontal cortex sub-regions have been found to signal key variables relevant to economic decision-making. These neural representations appear critical for generating an internal model of varying choice options and implementing goal-directed actions based on existing preferences, expectations, and current needs (see Rangel et al., 2008[97]).

1.6 Widespread and heterogenous representation of decision variables across frontal cortical regions

Taken together, the work described above indicates that the activity of individual neurons across the frontal cortex reflects a range of sensory, behavioural, and cognitive variables relevant to economic decision-making. However, there is still a considerable gap in our understanding of how these neural representations arise and are integrated to support decision-making behaviour.

One of the key challenges with addressing this gap is the heterogeneity of neuronal responses across the frontal cortex. While single-neuron studies like those described above have pointed to representational and functional differences between frontal sub-regions, it remains the case that populations of neurons across these regions

have been observed to be responsive to nearly all combinations of task-related variables[98;99]. For instance, while the study from Kennerley et al (2009) emphasised the over-representation and multiplexing of reward probability, magnitude, and cost in ACC, consistent encoding of the same variables was still observed across the other recorded regions (OFC/LPFC), albeit to a lesser extent. Indeed, value-, risk-, and reward-related signals have been observed across many regions of the frontal cortex - and indeed other cortical regions - in various behavioural contexts[78;85;90;100–104]. This suggests that neural representations of particular features of decision-making may not be restricted to specific regions of the frontal cortex, but rather that different regions may display under- or over-representations of certain decision variables. Such widespread and overlapping encoding of decision variables complicates efforts to disentangle the contribution of different sub-regions to distinct features of decision-making, as it suggests that the neural encoding in one sub-region only becomes meaningful when considered in the context of encoding in other regions.

Previous studies seeking to understand the encoding of decision variables in the frontal cortex have typically focused on a handful of neurons in one or a small number of brain regions in isolation. This is, in part, due to methodological limitations, as tools for monitoring neuronal activity have previously been limited to a small number of neurons in a spatially restricted area. While such approaches have provided valuable insight into the diverse array of information encoded in frontal neurons, these analyses are nonetheless prone to confirmation bias as it is impossible to compare the representations observed in the target region to those in unstudied areas. Furthermore, these studies have often utilised a wide variety of behavioural paradigms: from simple, head-fixed Pavlovian tasks[105] to complex, freely moving behaviour[92]. Given neural representations in the frontal cortex depend so heavily on task context and current goals[71;106], the lack of consistency in the behavioural tasks used makes it challenging to generalise findings across studies. Thus, to properly examine the organisation of decision variable representation across the frontal cortex, it is necessary to monitor the activity of populations of neurons across its constituent structures simultaneously within the same behavioural context. Thankfully, recent advancements

in the methods available for monitoring and manipulating neural activity in vivo - particularly in rodents - are opening up novel opportunities to study the neural basis of behaviour on a wider scale and with greater specificity than ever before[107–109]. The time is ripe, therefore, for large-scale experiments that seek to elucidate the neural mechanisms underpinning decision-making by monitoring the activity of large populations of neurons across multiple structures within the frontal cortex in the same behavioural context.

This will be the focus of the first two results chapters of this thesis: combining a decision-making paradigm in mice (Chapter 3) with large-scale electrophysiological recordings across the frontal cortex (Chapter 4) to examine and compare the encoding of economic variables in different frontal sub-regions during decision-making behaviour.

1.7 Projection-defined signals in frontal cortex

The frontal cortex is densely interconnected with almost all regions of the brain, making it ideally placed to act as a global regulator of brain function[45;49;110]. The development of electrophysiological and optical tools for targeting increasingly specific neural circuits has led to a growing appreciation of the role of different cell types in the mechanisms underlying complex behaviour[111;112]. In particular, evidence suggests that projection-defined populations of neurons may encode specific kinds of behaviourally-relevant information[105;113–115]. For example, an influential study by Otis et al. (2017)[105] used in vivo calcium imaging to examine how projection-defined dorsomedial PFC neurons encode reward-related information and how this encoding evolves during learning. The authors observed that frontal inputs to the nucleus accumbens (NAc) in the striatum developed excitatory responses to reward predictive cues over time, while inputs to the paraventricular thalamus (PVT) developed inhibitory responses. Optogenetically manipulating these pathways induced opposing effects on reward-seeking behaviour, with corticostriatal activation enhancing reward-seeking behaviour and corticothalamic activation suppressing it. In addition to high-

lighting specific and opposing roles for the PFC-NAc and PFC-PVT pathways in the dynamic control of reward-seeking behaviour, this study indicated that the encoding of behaviourally relevant information in PFC neurons may map onto specific inputs to downstream regions.

Similarly, Hirokawa et al.(2019)[115] used electrophysiological recordings in rats performing a perceptual decision-making task to explore the encoding of decision variables in orbitofrontal cortex (OFC). Employing a spectral clustering approach, they identified eight distinct clusters of OFC neurons that encoded key decision variables such as decision confidence, anticipated reward size, and integrated value. Subsequently, an in vivo optotagging approach was implemented to identify lateral OFC neurons projecting specifically to the ventrolateral striatum. The majority of these striatum-projecting OFC neurons were found to encode integrated value, displaying similar response profiles to only one of the eight identified OFC clusters. Thus, despite observing heterogeneous encoding of several decision variables at a population level, these authors found OFC neurons projecting to the striatum were more specific in the information they represented.

The observation that projection-defined neurons encode consistent kinds of information indicates that frontal neurons may play a key role in shaping the flow of information to downstream structures. This perhaps makes intuitive sense, as long-range projection neurons with relatively fixed representational content may increase computational efficiency by only providing downstream structures with information relevant to the computations they perform. In any case, these observations suggest that the underlying structure of the heterogeneous encoding of decision variables observed across the frontal cortex may be parsed out, at least in part, by a better understanding of the activity of projection-defined populations of neurons in an appropriate task paradigm. Progress in this regard may also help with efforts to establish targets for the treatment of psychiatric disease, as animal models of psychiatric disorders have already begun to demonstrate that pathological behavioural phenotypes can be regulated by manipulating specific projection-defined circuits. For instance, optogenetic stimulation of striatum-projecting OFC neurons was found to

suppress compulsive grooming in a mouse model of OCD-like behaviour[116]. However, the nature of information encoding in specific projection pathways is only starting to be addressed.

1.8 Frontal inputs to striatum and claustrum

Recent work on the 'projectome' of frontal cortex has identified the dorsal striatum and claustrum as two major targets of frontal input[48;117;118].

While the striatum has historically been considered a critical structure for volitional motor planning and execution, recent work has also implicated the striatum - particularly the dorsal striatum - in reward processing and executive function. Studies examining the role of the dorsal striatum in reward-related behaviours have highlighted contrasting roles of medial and lateral subregions, with dorsomedial striatum (DMS) and dorsolateral striatum (DMS) mediating goal-directed and habitual behaviours respectively[119–123]. Yin, Knowlton and Balleine (2005)[119] provided direct evidence for the role of the DMS in goal-directed action by lesioning the DMS of rats performing a lever-pressing task in exchange for food rewards. Rats with DMS lesions showed significant impairments in goal-directed behaviour: they continued to perform the level-pressing task even when rewards were devalued, indicating a loss of sensitivity to outcome value. Subsequent studies[124;125] indicated this abolishing of goal-directed learning occurs following both pre-taining and post-training lesions: suggesting that the DMS plays a key role in both the acquisition and maintenance of adaptive, goal-directed behaviour. This is supported by functional imaging studies in humans which have reported changes in BOLD signals in the DMS during anticipation of reward[126] and that correlate with performance in reversal learning and task-switching paradigms[127;128]. Furthermore, single neurons in DMS have been found to signal reward magnitude, probability, and value[122;129]. Together, these observations indicate the DMS plays a critical role in decision-making by representing associations between responses and outcomes and utilising those representations to support cognitive flexibility.

The claustrum, a thin sheet of grey matter nestled between the insula cortex and the ventral striatum, has also been identified as a major target of frontal cortical input. The claustrum has been a longstanding source of scientific intrigue due to its widespread connectivity with almost every region of the cortex, displaying the highest connectivity per unit volume of any region in the mammalian brain[130]. This striking connectivity has led to many hypotheses about the claustrum's function, ranging from multi-sensory integration and perceptual binding[131;132] to consciousness[133]. However, none of these early hypotheses garnered enough evidence to dominate: complicated by the limited tools available to study such a small and inaccessible structure. Notably, the majority of the claustrum's cortical connectivity extends to and originates from frontal cortical regions[117;134], suggesting that the claustrum may play a role in higher-order cognitive functions. Indeed, the claustrum has been found to modulate frontal cortical activity through its interaction with local inhibitory networks in the frontal cortex[135;136]. Additionally, a growing body of evidence has implicated the claustrum in cognitive processes such as attention control[137;138], impulsivity[139], engagement[140;141], salience and reward processing[142;143] and goal-directed action[144–146]. Together, these observations point to a potential role for the claustrum in goal-directed attention and cognitive control[147].

Critically, the dorsomedial striatum and the claustrum receive dense and integrative input from various frontal regions but are not themselves connected: indicating they are likely involved in different aspects of higher-order cognition. This makes these structures good candidates for examining the extent to which the encoding of decision-relevant information maps onto specific frontal projections. Thus, examining the activity of frontal neurons projecting to the dorsomedial striatum and the claustrum in the context of decision-making may provide new insight into the structure of information encoding in specific frontal circuits, and perhaps enable us to develop a better understanding of the role of corticostriatal and corticoclaustral circuits specifically in decision-making behaviour.

This will be the focus of the final chapter (Chapter 5): combining in vivo optotagging with large-scale electrophysiological recordings in mice performing a decision-making

task to investigate and compare the encoding of task-relevant information in dorsomedial striatum and claustrum-projecting frontal neurons.

1.9 Thesis overview

The key question this thesis seeks to answer is: how are the neural correlates of economic decision-making structured across multiple regions and projections of the frontal cortex? To address this, this project will use a combination of behavioural, electrophysiological, and optogenetic approaches to examine representations of decision variables within both region- and projection-defined populations of neurons in the frontal cortex. The first results chapter of this thesis (Chapter 3) will outline the behavioural task used to study decision-making: a value-guided decision-making task in which reward probability and reward magnitude are systematically manipulated to provide a means of examining the neural representation of these fundamental decision variables in the frontal cortex. This chapter will outline the key behavioural results indicating that this task offers a valid means of probing value-based decision-making behaviour in mice. The second results chapter (Chapter 4) will then cover experiments in which large-scale electrophysiological recordings were performed across frontal cortical regions in mice trained to an expert level in this task. By recording from a large number of neurons across multiple frontal cortical structures in the same task, this experiment enabled neural representations of decision variables to be directly compared across regions. The third and final results chapter of this thesis (Chapter 5) will examine the activity of projection-defined frontal cortical neurons, specifically those projecting to the dorsomedial striatum and claustrum, in the context of decision-making behaviour, and explore how encoding in these projection-defined neurons relates to the structure of neural representation observed at a population level across frontal regions. Overall, this thesis seeks to bridge the gap between population and single-cell approaches to studying the neural basis of complex behaviour and provide novel insight into the structure of frontal cortical representations underpinning economic decision-making in mice.

Chapter 2

General methods and materials

This chapter summarises the methods and materials common to most or all of the subsequent experimental chapters. Additional details or descriptions of experiment-specific methods will be described in the relevant chapters.

2.1 Subjects

All experimental procedures involving animals were conducted in accordance with the UK Animals in Scientific Procedures Act (1986).

The subjects used for all experiments in this thesis were male C57BL/6J mice, supplied by Charles River. Mice were aged 6-12 weeks at the time of their first surgery. All cohorts were group-housed with food and water available ad libitum, except for those used in behavioural experiments, which were water-restricted. Water-restricted mice had their body weight carefully monitored and maintained above 80% of their ad libitum level for the duration of training and recording. Mice were housed in a temperature-controlled (18°C - 22°C) and humidity-controlled (50%) facility with a 12-hour light/dark cycle.

A summary of the cohorts of mice used in each chapter can be found in **Table 2.1**.

Cohort	Purpose	Chapter
Cohort 1 (n= 23)	Behavioural training - economic decision-making task	Ch. 3
	Frontal cortex recordings during decision-making behaviour	Ch. 4
Cohort 2 (n = 18)	Optimising surgical procedures (intracranial stereotaxic injections, optic fiber implantations, head-plate surgeries)	Appendix B
Cohort 3 (n= 6)	Establishing an in vivo optotagging protocol to target projection-defined frontal cortical neurons	Appendix B
	Developing an analysis pipeline for electrophysiological and optotagging data	
Cohort 4 (n= 12)	Behavioural training - economic-decision making task	Ch. 3
	Frontal cortex recordings during decision-making behaviour	Ch. 4
	Projection-specific optotagging during decision-making behaviour	Ch. 5
Cohort 5 (n = 4)	Anatomical tracing of corticostriatal and corticoclaustral circuits	Appendix A

Table 2.1: Subject overview
An overview of the cohorts of mice used for each experiment in this thesis

2.2 Stereotaxic surgery

2.2.1 General surgical preparation and post-operative care

All surgical procedures were performed under isoflurane anaesthesia (3% in oxygen for induction, 1% - 2% in oxygen for maintenance). Animals' breathing rate and depth were monitored regularly throughout each procedure and the anaesthesia adjusted accordingly.

All animals received perioperative injections of 0.1 mg/kg buprenorphine (Vetergesic) and 5 mg/kg meloxicam (Metacam) to reduce post-operative pain. Their eyes were protected by regular application of artificial tears (Viscotears) to prevent drying. Aseptic technique was achieved through the use of sterile consumables and maintaining a sterile field.

After surgery, mice were placed in a warm recovery cage with analgesic jelly, mash, and additional food. Additional saline injections were given at the end of long surgical procedures (>2 - 3 hours) to reduce the risk of dehydration. Mice were weighed immediately after surgery and monitored closely for 5 days post-surgery to ensure recovery progressed as expected.

These pre-, peri-, and post-operative steps apply to all surgical procedures described below.

2.2.2 Head-plate surgery

Mice were implanted with custom-made titanium headplates to enable head-fixation during training and electrophysiological recordings. Following the general surgical preparation described above, the animal's scalp was sterilised with isopropyl alcohol (ChloraPrep) before being removed bilaterally. One drop of an acid primer (Super-Bond) was used to remove the periosteum from the skull's surface and induce a rough surface to promote good adhesion between the bone and headplate. Lambda

and bregma DV coordinates were used to ensure the skull was level. Adhesive dental cement (Super-Bond C&B, Sun-Medical) was used to implant the headplate directly above lambda. Marks were made anterior to bregma to indicate the location of frontal cortex for future procedures. Clear cement was then applied to the remaining exposed areas of skull and allowed to dry. Mice were left to recover for 5 days after surgery before handling and behavioural training began.

2.2.3 Intracranial injections and optic fiber implantations

In preparation for optotagging experiments, mice were injected with 150nl of an adeno-associated viral vector expressing Chronos, a high-speed channelrhodopsin[148], and a GFP fluorescent tag (pAAV-Syn-Chronos-GFP, titre = 1×10^{13} vg/mL) in their right frontal cortex. Initially, these injections consisted of a single injection of 150nl of virus at coordinates AP: +1.7 mm, ML: +0.6 mm, DV: -1.8 mm. These coordinates were later adjusted to three injections of 50nl at coordinates AP: +1.7 mm, ML: +0.6 mm, DV: [-1.8 mm, -2.2 mm and -2.6 mm] to increase the dorsoventral extent of opsin expression across frontal cortex. A small craniotomy was performed at these coordinates using a dental drill (NSK UK Ltd.) before injection to enable the pipette to enter the brain. Anterior-posterior and medio-lateral coordinates were measured relative to bregma and all dorsal-ventral coordinates from the surface of the brain.

Optic fibers (200 µm diameter, 3.5mm length, 0.37 NA, Neurophotometrics) were chronically implanted above the dorsomedial striatum and/or the claustrum in the same hemisphere (right) as the viral injection. A small craniotomy was performed at these coordinates before implantation to permit the insertion of the optic fiber. The coordinates used for the implantation of the optic fiber above CLA were AP: +1.3 mm, ML: +2.8 mm, DV: -2.2 mm. The coordinates used for the DMS optic fiber were AP: +0 mm, ML: +1.2, DV: -2.4 mm and at an implantation angle of 17°. The angle of implantation was set at 17° to ensure sufficient space between the two fibers to permit both to be connected to optic patch cords simultaneously during recordings.

2.2.4 Tracing surgeries

Mice used for anatomical projection tracing were injected with 250nL of an adeno-associated viral vector expressing a Cre-dependent tdTomato fluorescent tag (pAAV-Syn-FLEX-rc[ChrimsonR-tdTomato]) in the right frontal cortex and 250nL of retro-Cre (AAV.hSyn.Cre.WPRE.hGH) in the DMS. Injections in the frontal cortex were made in the right hemisphere at coordinates AP: +1.7mm, ML: +0.6mm, DV: [-1.8mm, -2.2mm, -2.6mm]. DMS injections were made at coordinates AP: +0.7mm, ML: +1.2mm, DV: -2.5mm. Mice were left for 5 weeks after surgery to permit adequate time for viral expression before they were sacrificed for imaging.

2.3 Behaviour

2.3.1 Materials and apparatus

The behavioural rig consisted of a custom-made head-fixation set-up located a fixed distance (11cm) away from a small screen (LG display - 9.7" diagonal, Adafruits). Head-fixed mice were positioned in front of this screen with their forepaws on a small rotating wheel. The wheel was connected to a rotary encoder (Kubler Incremental Encoder), which recorded precise measurements of wheel velocity and direction. A plastic water delivery spout connected to a solenoid was placed in front of the mouth to permit delivery of water rewards. Water delivery was calibrated every four months to ensure water rewards matched the predicted volumes. Body and eye cameras were used to track pupil diameter, licking, and task-related movement during recordings. The behavioural task was controlled by custom software written in MATLAB.

2.3.2 Visual economic decision-making task

Mice were trained to perform a novel visually-guided decision-making task. This task was based on an established visual detection task[149] adapted to probe economic

decision-making through the introduction of visual stimuli corresponding to different reward magnitudes and probabilities.

On each trial, an abstract visual stimulus was presented on the right or left-hand side of the screen (±35° azimuth, 0° altitude). The visual stimulus remained in a fixed position for a brief interactive delay period [0.2 - 0.3s]. Mice were then able to turn the wheel left or right to 'select' the stimulus by moving it to the centre of the screen. A correct selection prompted the delivery of the corresponding water reward. Moving the stimulus to the far side of the screen was considered an incorrect selection and resulted in no reward. There were three possible visual stimuli, each corresponding to a specific value and probability of reward: a large, guaranteed reward condition (3 µL, 100% of correct trials rewarded), a small, guaranteed reward condition (1.5 µL, 100% of correct trials rewarded) and a gamble condition (3 µL, 50% of correct trials rewarded). These stimuli will be referred to as 'Large', 'Small', and 'Gamble' moving forward. The correspondence between abstract visual stimuli and trial condition was randomised across mice. All stimuli were presented in a pseudorandomised order throughout the session. The inter-trial interval varied randomly from 1 - 3s. Before the start of each trial, there was a variable time window [0.2 - 0.5s] in which no movement of the wheel was permitted.

The task was stopped when the mouse lost motivation, indicated by a slowing of reaction time, or had completed sufficient trials to meet its daily water requirement (600 - 1000µl, adjusted for current body weight).

2.3.3 Behavioural training

To perform the task described above, mice were trained for approximately half an hour a day for 4-8 weeks. Mice were handled and acclimatised to the behavioural rig for 3 days before training to reduce stress during head-fixation. In the first stage of training, mice were only presented with the visual stimulus corresponding to the large reward (3 µL, 100% of correct trials rewarded) on either the right or left side of the screen each trial. Once performing reliably above chance (>60 % correct on

either side) on this trial type, the gain of the wheel was reduced to increase the size of the movement required to select a stimulus. Additional stimuli predicting different outcomes (Small and Gamble stimuli) were then added in series when mice reached an accuracy threshold of >70% for each stimulus type and maintained this for 4 - 5 days. During training, incorrect choices triggered a repeat of the previous trial. Mice were deemed ready for electrophysiological recording from when they reached expert performance on this task: defined as displaying >80% accuracy across all trial types.

2.3.4 Behavioural inclusion criteria

A set of quality control criteria was used to define which behavioural sessions would be included in the final behavioural dataset. Behavioural sessions were filtered based on payoff values, as some of the earliest animals trained in this task were presented with different stimulus-reward contingencies. The sessions retained had a minimum payoff of either 1.4 or 1.5 and a maximum of 2.8 or 3.0. Trials were also excluded if they were repeat trials or displayed response times exceeding 10 seconds. The initial two sessions when mice were transferred between rigs were also omitted to account for potential equipment calibration issues and to allow mice to adjust to the new experimental environment. Sessions that did not showcase all three stimulus conditions in at least two preceding sessions, ensuring a comprehensive exposure to stimuli, were also excluded. Lastly, sessions were discarded if they displayed less than 80% behavioural choice accuracy across all trial types.

2.4 Large-scale electrophysiological recordings

2.4.1 Electrophysiological recordings

Neuronal activity in the frontal cortex was recorded using high-density silicon probes (Neuropixels[150]). A craniotomy surgery was performed the day before the first acute recording to provide a route for probe implantation. A small 3D-printed, semi-circular

head cone was also attached to the skull using dental cement. The purpose of this cone was to create a saline bath over the craniotomy to keep the exposed brain tissue clean and moist for the duration of the recording. The saline bath was also connected to a grounding electrode to ensure the recording area was electrically isolated. The craniotomy was covered in a protective silicon material (Kwik Cast) and the animal was left to recover overnight.

On recording days, mice were head-fixed in the recording rig. The protective material covering the craniotomy was removed and the head cone was filled with saline. The grounding electrode was connected to the Neuropixels probe and placed in the saline bath. To verify the location of the probe tract during post-hoc histology, probes were coated in a lipophilic red-fluorescent dye (Invitrogen, Vybrant Dil Cell-Labeling Solution, V22885) before recording. The probe was then inserted at a low speed through the dura and into the brain using a manipulator (Sensapex Micromanipulator). Once at the target depth, the probe was retracted by 100µm to release tension within the tissue and left in place for 5 minutes before recording began. Electrophysiological recordings were performed using SpikeGLX software. The behavioural task was initiated just after the recording was started and ran for between 20 and 40 minutes.

Recordings were performed on consecutive days for an average of 7 ± 2 days for each mouse.

2.4.2 Electrophysiological data processing

Recordings were performed using SpikeGLX software. Electrophysiological data was sampled at 30 kHz and subsequently processed using the automatic spike sorting software KiloSort2[151]. To ensure only well-isolated single units were included in subsequent analysis, spike-sorted data was then manually curated using Phy2 to distinguish between single and multi-unit clusters and remove noise. A "good" single unit was defined as one displaying clear waveforms, a valley-shaped correlogram, $< 0.2\%$ refractory period violations, and an attenuation in spike amplitude with increasing distance from the original recording channel (see Fig 2.1).

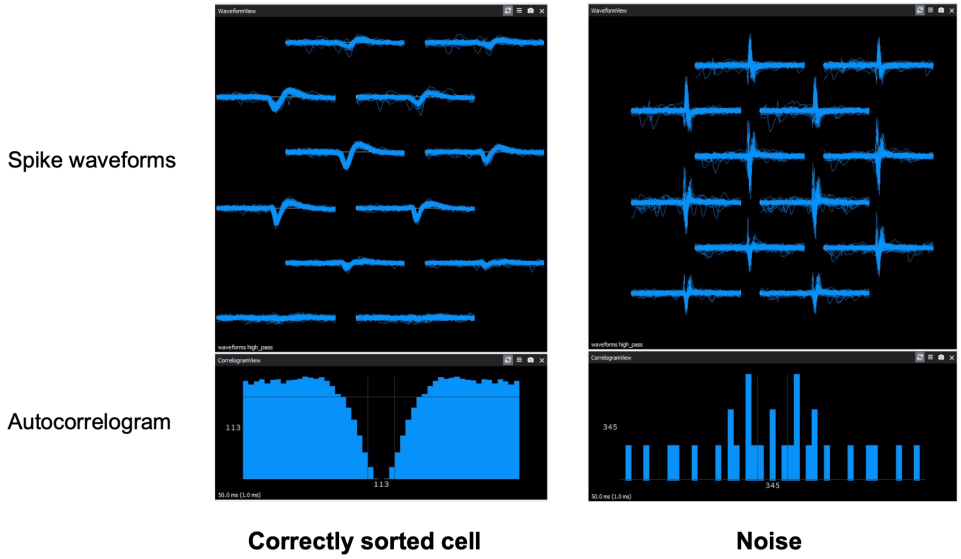


Figure 2.1: Manual curation of spikesorted data was used to separate 'good' single units from multi-unit activity and remove noise. Noise may be differentiated from real spiking activity as it does not produce a clear waveform, lacks a distinct inter-spike interval, and shows no attenuation in spike amplitude as the distance from the original recording channel increases (left:a correctly identified single unit with a clear waveform and inter-spike interval, right: noise)

2.5 Optotagging projection-defined neurons

2.5.1 Optotagging protocol

In a subset of animals (n = 12), an optotagging protocol was implemented immediately after the behavioural task, but as part of the same electrophysiological recording, to identify frontal cortical neurons projecting specifically to dorosomedial striatum and/or claustrum. This was achieved by delivering a series of laser pulses (VersaLase, Laser 2000) via the optic fibers implanted above DMS and CLA to induce antidromic activation of opsin-containing frontal neurons projecting to these downstream structures. The CLA and DMS stimulation blocks occurred in series. On each trial within a stimulation block, mice either received a drop of water via a water delivery tube (1µL) or a laser pulse (wavelength = 470nm, duration = 1ms, laser power = ˜50mW) via the optic fiber. Stimulation blocks consisted of 50, 150, or 500 laser pulse trials

with water delivery trials randomly interleaved. Water delivery trials were required to accurately align the electrophysiological data with the behavioural time-base. The intertrial interval varied from 4 - 6s to permit sufficient time for the network to return to baseline between water delivery and laser pulse trials. Electrophysiological recordings were performed and processed in the same manner as described above.

Five recording sessions were discarded due to technical faults. An additional session was discarded due to the observation that no neurons were found to be task-responsive. Finally, the CLA-stimulation data collected from three mice were discarded due to incorrectly targeted optic fibers or Neuropixel implantation trajectories, as confirmed by histological examinations. The final optotagging dataset consisted of 46 recording sessions across 12 mice.

2.5.2 Optotagged neuronal selection criteria and validation

DMS- and CLA-projecting neurons were identified using a set of criteria consistent with existing optotagging literature[17;113–115]. First, all optotagged units were required to be well-isolated single units following Phy curation of Kilosorted data. A 'good' unit was defined as one displaying clear waveforms, a distinct inter-spike interval with few refractory period violations, and an attenuation in spike amplitude with increasing distance from the original recording channel (see above). Secondly, optotagged units were required to have a significant stimulus-aligned latency test (SALT, $p < 0.05$). Based on analysis from Kvitsiani et al (2013)[111], SALT is a statistical test that compares the distribution of first spike latencies in a 10ms window after laser delivery with multiple 10ms epochs in a stimulus-free baseline period. Only units with a significant SALT statistic were included in subsequent analysis, indicating their firing rate following laser onset was significantly different to their baseline activity. Finally, optotagged units were required to display short latency and high fidelity laser responses, displaying laser-evoked spikes in a 10ms window after laser-onset on >40% of laser pulse trials (Fig 2.2A), and have a significant correlation ($r > 0.9$) between their spontaneous and laser-evoked waveforms (Fig 2.2D). Implementing these

criteria identified a subset of DMS- and CLA-projecting neurons, forming seperable clusters relative to all recorded neurons (Fig 2.2B-C) with a high correlation between spontaneous and laser-evoked waveforms (Fig 2.2E-F).

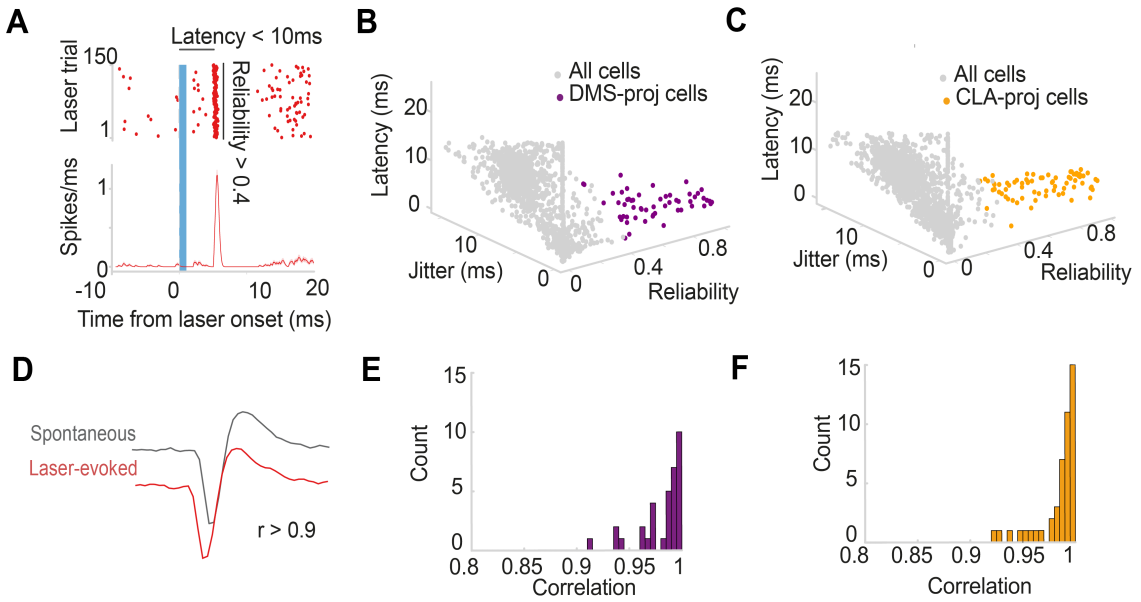


Figure 2.2: In vivo optotagging criteria for identifying CLA- and DMS-projecting frontal cortical neurons A) To be deemed 'optotagged', neurons were required to a mean first spike latency $<$ 10ms and display a laser-evoked spike in this window on more than 40% of laser pulse trials. Their post-laser firing rate also had to be significantly different to baseline (stimulus-aligned latency test, $p < 0.05$) B) Reliability, latency, and jitter of post-laser activity for all neurons in DMS-stimulation blocks, neurons passing the optotagging criteria indicated in purple C) Same for CLA-stimulation blocks, neurons passing the optotagging criteria indicated in orange D) Optotagged neurons had to display a high correlation ($r > 0.9$) between spontaneous and laser evoked waveforms E) Distribution of waveform correlations for final dataset of DMS-projecting neurons (see Chapter 5) F) Same for CLA-projecting neurons (see Chapter 5)

2.6 Histology and probe alignment

2.6.1 Histological processing

For post-hoc histology, mice were terminally anaesthetised and transcardially perfused with 4% PFA in 0.1M PBS. The brains were extracted and post-fixed in the same

solution overnight at 4 °C. They were then sliced into coronal sections (100µm thick) using a Vibratome (Leica Microsystems). The slices were washed in 0.01M PBS, followed by 1µg/mL DAPI, and placed on a shaker for five minutes. The slices were then mounted on gelatin-coated slides and cover-slipped with a mounting medium. Slides were left to dry overnight before being imaged on a fluorescence microscope. Fluorescence imaging was carried out using a Leica fluorescence microscope. Images were taken at 1.6X and 10X at 365nm and 530nm to visualise the DAPI stain, opsin fluorescence, optic fiber tracts, and the DiI probe tracts.

2.6.2 Probe tract alignment

To verify the location of the Neuropixels probe tract, probes were coated in the lipophilic dye DiI prior to recording. Images of consecutive slices containing visible probe tracts were taken using a fluorescence microscope. These slices were then registered on the Allen Common Coordinate Framework (CCF)[152] and used to reconstruct a 3D probe tract. To ensure accurate alignment of slices to the Allen CCF, certain anatomical landmarks were used to guide the alignment including the corpus callosum, anterior commissure, and the midline. Ventricles were not used as ventricle size can be affected by perfusion and variation between mice[153]. Mapping the reconstructed probe tract to the IBL Electrophysiology Atlas was then used to both verify the probe was targeted correctly and map each kilosorted unit to its corresponding brain region.

2.6.3 Anatomical tracing image analysis

In order to quantify the relative axonal expression of TdTomato in DMS and CLA for projection tracing, we first established the boundaries of these regions of interest (ROIs) by mapping the imaged brain slices to the Allen CCF Brain Atlas[152]. The resulting images were then processed and quantified using ImageJ.

To examine whether opsin-expressing DMS-projecting neurons sent axon collaterals

to CLA, we quantified the fluorescence in CLA relative to both the fluorescence at the injection site in the DMS and the background fluorescence. The purpose of these comparisons was to account for differences in viral expression between mice and to ensure the fluorescence observed in CLA was significantly different from non-specific background fluorescence. ImageJ was used to manually select the ROIs according to the aligned Allen Brain Atlas. The mean grey value and the integrated density were then extracted for the DMS, ipsilateral CLA, and contralateral CLA. Five background measurements were also taken and averaged. The percentage fluorescence in each area was calculated as a proportion of DMS fluorescence. Paired t-tests were then conducted to test if the differences in fluorescence between the ipsilateral CLA (ipsCLA) and background, the contralateral CLA (contCLA) and background, and the ipsCLA and contCLA were statistically significant.

2.7 Data analysis

Statistical analyses were performed using custom scripts written in MATLAB, R, and Python. We did not use statistical methods to determine sample size and no randomisation or blinding was used. Statistical tests were two-tailed and a p-value of 0.05 was used as a threshold for significance throughout unless otherwise stated.

2.7.1 Population decoding

A linear population decoding pipeline was implemented to examine the extent to which trial conditions could be decoded from the population activity of task-responsive neurons. Neurons were deemed to be responsive if they displayed a significant difference from their pre-trial baseline activity in defined windows around key task events using a Bonferroni correct Wilcoxon signed-rank test (see Appendix A.2 - Identifying task-responsive neurons). This and all subsequent analyses only considered trials with a correct behavioural response, ensuring that mice were engaged with the decision-making task.

Kernel regression

Electrophysiological data from high-quality recording sessions was initially fit with a kernel regression model to remove the effects of general stimulus and action signals, not specific to the individual task conditions. To this end, a model with a single stimulus kernel and a single action kernel was fit to the trial-by-trial responses of each neuron[154;155]. While this model was able to capture average stimulus and action-related activity, it could not, by design, capture differences in neural activity driven by individual conditions (i.e. Large, Small and Gamble trials). Therefore, subtracting cross-validated model predictions from the raw data retained condition-specific neural activity, while removing common stimulus- and action-related activity.

In order to leverage the large number of recorded neurons for population decoding, classifiers were trained on the combined data across all recording sessions. Sessions were combined by randomly subsampling each session's trials to form sets of 30 Large, Small and Gamble trials each, counterbalancing left and right stimulus presentations. Trials in each respective condition were then paired across sessions, to form a large surrogate recording with 30 trials per condition. To mitigate variability introduced by this random sampling and matching, this procedure was repeated 100 times and decoding results were averaged across all 100 iterations.

Time-resolved decoding

The first decoding analysis was aimed at identifying a time window relative to stimulus presentation during which differences between Large, Medium and Gamble conditions could be reliably decoded from recorded neurons, irrespective of brain region. To this end, a three-class linear Support Vector Machine (SVM) was trained and tested to classify the three condition labels (Large, Small, Gamble) based on binned spiking activity in 10 ms time windows ranging from 50 ms before to 350 ms after stimulus onset using 10-fold cross-validation. Performing this procedure for all combinations of training and testing time points yielded a temporal generalisation matrix[156], revealing

the time course and dynamics of the neural representations differentiating between trial conditions. To statistically assess decoding performance, a cluster-based permutation test was performed[157], by randomly shuffling trial labels and repeating the above decoding procedure. The size of the largest spatial cluster in the temporal generalisation matrix was recorded for each of 1,000 permutation iterations. Spatial clusters were defined as contiguous training-testing bins exceeding 5% decoding accuracy above chance level (1/3, cluster-forming threshold). The size of a cluster was defined as the summed decoding accuracy across all training-testing bins in the cluster. The size of the largest empirical cluster was compared to the permutation distribution, and a p-value was calculated as the proportion of permutations that resulted in a cluster that was equal or larger than the empirical cluster. The critical p-value was 0.05 (one-sided test).

Population and region-specific binary decoding

After establishing a post-stimulus time window during which trial conditions could be reliably decoded, binary classifiers were employed to investigate which pairs of conditions contributed to the decoding in this time window. Binary classifiers, distinguishing between all three combinations of conditions (Large vs Small, Large vs Gamble, Small vs Gamble) were implemented both for pooled activity across all recorded neurons and separately for each population of region-defined neurons (prelimbic, infralimbic, etc). Classifiers were trained and tested on time-averaged neural activity between 150 and 350 ms post-stimulus - the window in which time-resolved three-class decoding accuracy was highest. The decoding accuracy of each binary classifier was tested against chance using a permutation test. In this test, trial labels were randomly shuffled within each binary comparison, and the decoding analysis was repeated across 1,000 iterations. P-values were calculated as the proportion of permutations that yielded an equal or larger decoding accuracy than the one empirically observed decoding accuracy. The critical p-value was 0.05 (one-sided test). To statistically assess differences in decoding accuracy between binary decoders (e.g.

Large-Gamble versus Small-Gamble) a similar permutation test was used. For example, when testing Large-Gamble versus Small-Gamble classifiers, the trial labels between Large and Small trials were randomly shuffled and the decoding procedure was repeated with these shuffled trial labels. The empirical difference in decoding accuracy between Large-Gamble versus Small-Gamble was then compared to the permutation distribution of decoding accuracy differences. The critical p-value was 0.025 (two-sided test).

As the number of neurons in each brain region was variable, the above decoding analysis was repeated for sub-sampled populations of neurons in steps of 50 neurons (up to a maximum of 800 neurons) to compare decoding accuracy across brain regions at comparable neuronal population sizes. Each sub-sampling was repeated 100 times, to average out variability introduced by drawing random samples of neurons.

Spatial binary decoding

To investigate the spatial organisation of neural representations of economic variables beyond brain region labels, the spatial distribution of neural activation weights was examined along the mediolateral (ML) and dorsoventral (DV) axes of the frontal cortex. Neural activation weights for EV and Risk were derived from binary classifiers distinguishing Large versus Small (different EV), and Gamble versus Small (different Risk), respectively. Activation weights for each neuron were computed by left-multiplying the SVM weights by the data covariance matrix[158]. Neurons were grouped into spatial bins (ML spatial bin size = 0.1mm, DV spatial bin size = 0.2mm) containing $>$ 50 neurons and averaged to determine the mean absolute activation weight for each bin. The change in average activation weights across the ML and DV axes was then tested by fitting a simple linear regression to the distribution of average activation weights across ML and DV bins and extracting slope and intercept estimates. Permutation tests with shuffled coordinate labels, and shuffled activation weights between comparisons, were then used to test that these empirical activation gradients across trial conditions were significantly different from zero and each other. The critical

p-value was 0.025 (two-sided test).

2.7.2 Single cell analysis

Single-cell responses to task events were examined by calculating the mean residual firing rate for each neuron in 200ms windows around stimulus onset and choice across trial conditions (Large, Small, Gamble). A Wilcoxon rank sum test was used to identify neurons whose activity was significantly different across pairs of conditions within these windows (Large v Small, Large v Gamble, Small v Gamble). Z-tests were then used to assess whether the proportions of neurons modulated by different trial conditions were significantly different across neuronal populations.

To visualise the activity of individual neurons across time or around key task events, the spiking activity of each neuron was grouped into 10ms bins, smoothed with a Gaussian filter, and plotted as a peri-stimulus time histogram.

Chapter 3

Establishing a economic decision-making task in mice

3.1 Introduction

The goal of this thesis is to investigate how the representations of economic variables are structured across the regions and projections of the frontal cortex. A critical initial step in this endeavour is, therefore, to establish a behavioural paradigm in which mice make decisions in response to stimuli predictive of different outcomes. This necessitates a task design that both engages frontal cortical circuits and captures essential characteristics of economic decision-making behaviour. Such a task must also provide an informative read-out of choice/preference that is robust to noise (i.e. random errors) and alternative strategies, such as visual or motor biases. Finally, it must be compatible with methods for monitoring neural activity such as large-scale electrophysiological recordings.

Two key variables commonly manipulated in economic decision-making paradigms are expected value (EV) and risk. Expected value refers to the probability-weighted average of a potential outcome. In other words, the return (e.g reward magnitude) multiplied by the likelihood of securing it (i.e reward probability). Behavioural economics has widely demonstrated that decision-making behaviour adapts to changes in these variables. Expected utility theory, for instance, posits that rational actors

make choices that maximise expected returns - indicating that decision-making behaviour is driven by a desire to secure higher value rewards[1;2;159]. Another influential model, prospect theory, suggests that people tend to prefer certain gains (low risk) over larger but less certain gains (high risk)[3] - indicating human decision-making is strongly influenced by an aversion to loss. While the predictions of these particular models of human decision-making behaviour are not always borne out experimentally[160;161], these influential theories nonetheless emphasise the significant influence expected value and risk can have on decision-making processes.

To probe economic decision-making behaviour in mice, we developed a novel two-alternative forced-choice (2AFC) visually-guided decision-making task. This task is based on an established perceptual decision-making framework[149] but modified to examine economic decision-making through the introduction of visual stimuli that signal different reward probabilities and magnitudes. This task utilises head-fixation to facilitate stable and high-quality neural recordings, a necessary attribute to permit the large-scale electrophysiological recording experiments described in subsequent chapters (Chapters 4 and 5).

While 2AFC tasks are becoming increasingly common in mouse perceptual decision-making studies, there are currently only a small number of studies using 2AFC tasks to study economic decision-making in rodents[115]. The 2AFC task structure presents several advantages over simpler paradigms commonly used to study economic decision-making behaviour in rodents. In a Go/No-Go (GNG) task, for instance, subjects are required to execute actions in response to certain stimuli (Go) and suppress actions in response to others (No-Go). While this approach is useful for studying response inhibition and executive control, it is vulnerable to fluctuations in an animal's motivation and response criterion[162]. In contrast, a 2AFC task provides symmetrical response options, ensuring that random choices result in an average accuracy of 50%. This feature eliminates the need to establish a baseline response rate, as is the case in a GNG paradigm, and enables reliable estimation of biases and lapse rates based on the resulting psychometric curves.

This task was also designed to bridge a key gap between primate and rodent decision-making literature. In rodent decision-making studies, animals are often presented with reward-associated stimuli whose sensory features directly relate to the features of the outcome they predict. For example, a tone whose frequency scales with reward probability[163;164] or different odour concentrations that scale with reward magnitude[165]. However, in such cases, the sensory features of the presented stimuli are highly correlated with the outcomes they represent: making it challenging to separate value signals from sensory-evoked activity. In contrast, primate decision-making paradigms often involve the presentation of abstract visual stimuli whose visual features are entirely independent of the associated outcome[80;85]. In an attempt to align this task with existing models of economic decision-making in primates, mice were trained to respond to abstract visual stimuli whose sensory features did not map onto the corresponding rewards. While the degree of abstraction demanded by this task likely increases the training time necessary to reach 'expert' performance relative to simpler tasks, it provides a valuable opportunity to investigate the neural basis of economic decision-making in a more translational manner.

This chapter will outline the behavioural results arising from the decision-making paradigm used in this study. The high level of choice accuracy observed across all stimulus conditions demonstrates that mice successfully learn this task. Critically, these data also reveal that mice are sensitive to the expected value of presented options, as their performance observed to scale with the expected value of the presented stimuli. This behaviour was found to be robust against sensorimotor biases and consistent across mice. Together these results demonstrate that this behavioural paradigm offers an appropriate means of probing the neural underpinnings of economic decision-making in mice.

3.1.1 Task outline

Mice were trained to perform a 2AFC visually-guided decision-making task in which reward size and probability were systematically manipulated across trials. In this

task, head-fixed mice were positioned in front of a small screen with their paws on a rotating wheel (Fig 3.1A-B). On each trial, a visual stimulus predicting a specific magnitude and probability of water reward appeared on the right- or left-hand side of the screen (Fig 3.1C). Mice were required to move the stimulus to the centre of the screen by turning the wheel in the correct direction, prompting the delivery of the corresponding water reward. Incorrect trials, in which the mouse turned the wheel in the wrong direction thereby moving the stimulus to the far edge of the screen, were not rewarded.

Three abstract visual stimuli predicting different magnitudes and probabilities of reward were used in this task (Fig 3.1D). These three stimuli will be referred to as 'Large', 'Small', and 'Gamble' stimuli. The 'Large' stimulus predicted a large reward (3 µL) delivered with 100% probability. The 'Small' stimulus predicted a small reward (1.5 µL), also delivered with 100% probability. The 'Large' stimulus thus had double the expected value of the 'Small' stimulus. Finally, the 'Gamble' stimulus predicted a large reward (3 µL), delivered with a probability of 50%, yielding a stimulus with the same expected value as the 'Small' stimulus but with higher risk. The term 'risk' is used here in an economic sense, referring to the probability of not receiving a reward rather than the presence of an aversive outcome.

The mapping between specific visual stimuli (cross, circle etc) and reward conditions was randomised across mice to ensure preference for specific task conditions was not the result of the visual features of the presented options (see Section 3.2.3 - Accounting for the effects of visual or motor biases on task performance).

A more detailed description of the behavioural task can be found in the General methods and materials (Section 2.3).

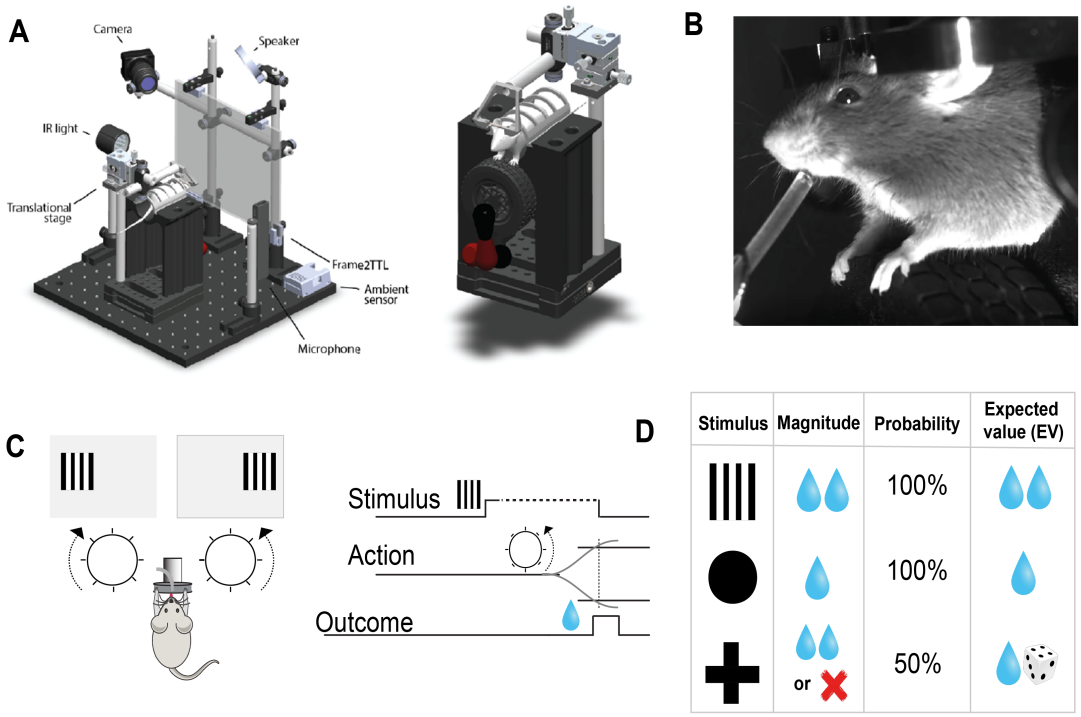


Figure 3.1: A visually-guided decision-making task to probe economic decision-making in mice A) Set-up of the behavioural rig B) Head-fixed mice were position in front of a water delivery spount with their forepaws on a rotating wheel C) Schematic of decision-making task. On each trial, a visual stimulus associated with a specific magnitude and probability of reward appeared on the right OR left side of the screen. Mice were required to turn the wheel to indicate their selection. Correct selections prompted the delivery of the corresponding reward D) Three different stimuli were used in this task. A Large stimulus (3 µL, 100% reward probability), a Small stimulus (1.5 µL, 100% reward probability), and a Gamble stimulus (3 µL, 50% reward probability). The mapping between specific visual stimuli (cross, circle etc) and reward conditions was switched randomly between mice.

3.2 Results

3.2.1 Task performance reflects differences in expected value

35 mice were trained to expert-level performance on this task. All mice displayed a high level of accuracy across stimulus conditions (Fig 3.2A, mean across all conditions = 89%), demonstrating they were able to relate the movement of the wheel to the selection of the on-screen stimulus and subsequent reward delivery. The ability to

turn the wheel to select the appropriate reward was acquired rapidly during training (see Section 3.2.4 - Task acquisition for details).

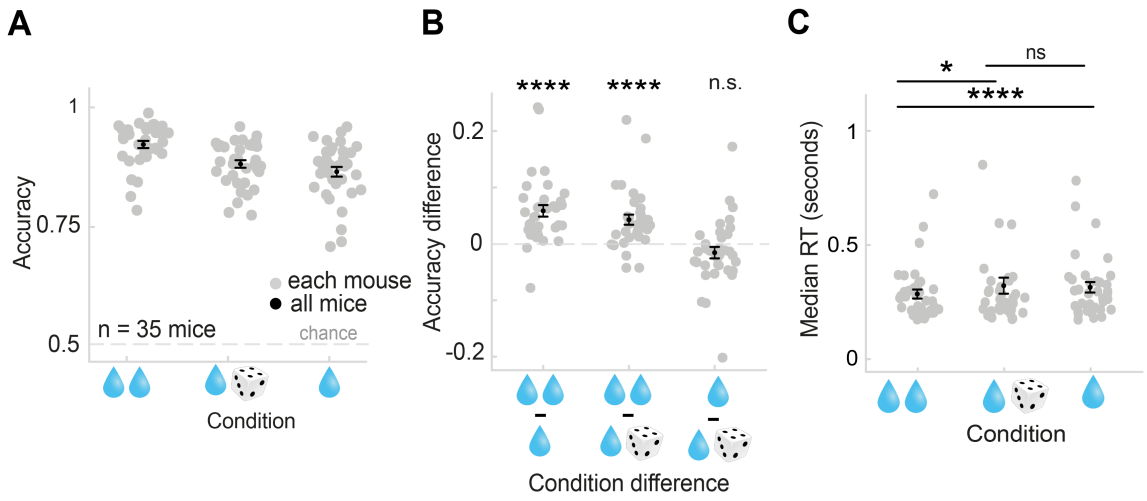


Figure 3.2: Choice accuracy scales with expected value (EV) A) Behavioural accuracy across trial conditions for all mice (n=35). Mice display high accuracy across all stimulus conditions. Grey dots correspond to individual mice, black dots to mean performance across mice (error bars = s.e.m). B) Difference in accuracy for pairs of stimulus conditions. Choice accuracy scales with expected value: mice display significantly greater performance for trials corresponding to a high EV (Large) compared to a low EV (Gamble or Small). The difference in performance for trial types with the same EV (Small or Gamble) is not significant. C) Response times (RT) scale with expected value. RTs are significantly lower for Large trials compared to both Gamble and Small trials

Comparing performance on different task conditions, a repeated measured ANOVA found a significant effect of stimulus condition on choice accuracy (two-way ANOVA, p ¡ 0.0001). Accuracy was greater for trials corresponding to the Large condition (mean accuracy = 92.3%, s.e.m = 0.75%) compared to the Gamble (mean accuracy = 88.2%, s.e.m = 0.79%) and Small conditions (mean accuracy = 86.5%, s.e.m = 1.02%). Post-hoc comparisons using pairwise t-tests (Fig 3.2B, Table 3.1) confirmed that mice displayed significantly greater accuracy for Large trials compared to both Small ($p < 0.0001$) and Gamble trials ($p < 0.0001$). There was no significant difference in performance on Small and Gamble trials ($p = 0.105$).

The Large stimulus has a higher expected value (3 µL) than the Small and Gamble stimuli, which have the same expected value (1.5 µL). The differences in accuracy ob-

Condition comparison	Mean Accuracy Difference	s.e.m	p-value
Large v Small	0.0571	0.0103	$3.86e^{-}6$, ****
Large v Gamble	0.0408	0.0080	$1.12e^{-}5$, ****
Small v Gamble	-0.0163	0.0100	0.105, ns

Table 3.1: Accuracy differences across trial types

served across stimulus conditions suggest that animals' are sensitive to these difference in expected value. If mice were not sensitive to these differences, we would expect performance across all stimulus conditions to be largely equal. This is consistent with previous literature, which has observed that rodents' performance on decision-making tasks is consistently higher in response to stimuli predictive of higher value outcomes[163;164;166]. These results indicate that mice are sensitive to the differences in the outcomes these abstract visual stimuli represent and are more motivated when responding to stimuli predicting larger rewards.

Taken together, these behavioural results indicate that mice can make visually-guided decisions based on abstract stimuli. Furthermore, the observed scaling of performance with the expected value of presented options suggests that mice are making value-based decisions in this task.

3.2.2 Response times reflect differences in expected value

Having established that animals' choices in this task reflected differences in expected value, their response times across conditions were then compared (Fig 3.2C). Response time and movement vigour have previously been used as proxy read-outs of motivational state and/or subjective economic utility[167]. Given the observed scaling of performance with expected value, indicating a preference for higher-value outcomes, it seemed plausible that mice with expert-level performance would also display faster response times in response to higher-value stimuli.

A repeated measures ANOVA found a significant effect of stimulus condition on median z-scored response times (two-way ANOVA, $p = 0.013$). Mice displayed a significantly lower z-scored response time for Large trials compared to both Small ($p <$

Condition comparison	Median response time (s)	sem
Large	0.285	0.0198
Gamble	0.305	0.0249
Small	0.314	0.0230

Table 3.2: Response times across stimulus conditions

Condition comparison	RT Difference	s.e.m	p-val
Large v Small	-0.0292	0.007	$1.49e^{-}5$, ****
Large v Gamble	-0.0195	0.0091	0.043, *
Small v Gamble	0.0098	0.0111	0.054, ns

Table 3.3: Response time reflects differences in expected value

0.0001) and Gamble trials (p = 0.043). No significant difference in response time was observed between Small and Gamble trials (p = 0.054). It is necessary to z-score this data as absolute response time can be affected by variability in the baseline response times across sessions and mice. Thus, response times were faster for stimuli corresponding to greater expected values: likely reflecting a heightened motivational state in this condition.

Two-stimulus trials

To provide further evidence that mice are sensitive to differences in the expected value of these stimuli, a separate study in the lab trained a subset of animals (n = 13) on 'two stimulus trials'. In this version of the task, mice were presented with two visual stimuli simultaneously and were required to choose between them. Consistent with their performance in single stimulus trials, animals presented with a choice between visual stimuli corresponding to different expected values consistently selected the higher value stimulus (Fig 3.3, Table 3.4) on both trial types (Large vs Small, t-test, p-value < 0.0001; Large vs Small, t-test, p-value < 0.0001). When offered a choice between two stimuli with the same EV (Small vs Gamble), mice did not display a significant preference for either stimulus (Small v Gamble, t-test, p = 0.55) - also consistent with the single stimulus task. The significantly greater preference for higher-value stimuli observed across mice when offered a choice between stimuli

predicting a high- or low-value outcome provides additional evidence that mice are sensitive to differences in expected value and can make economic decisions to maximise reward.

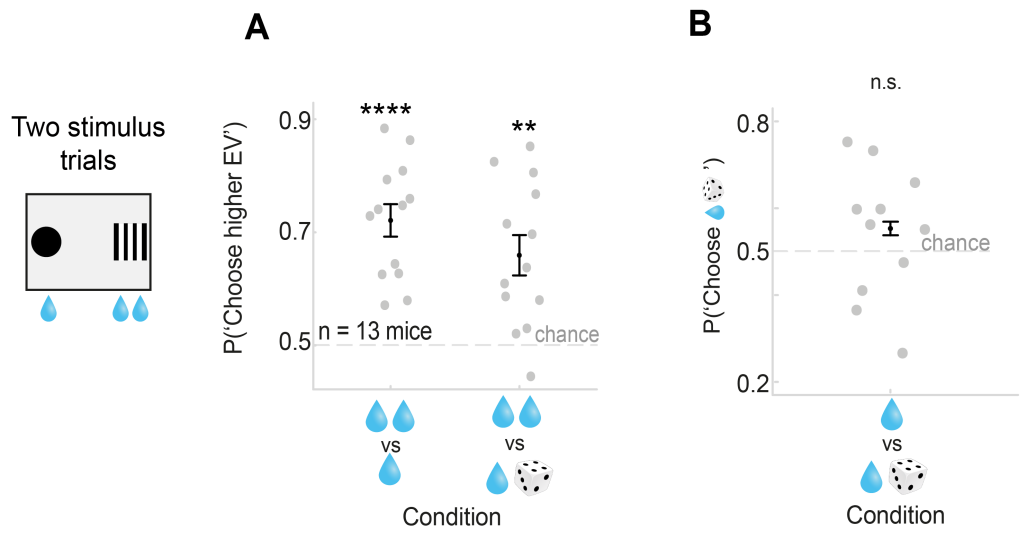


Figure 3.3: Performance on two stimulus trials A) Accuracy on two stimulus trials (n = 13 mice). Mice trained on two stimulus trials consistently selected the higher value stimulus when presented with options that differ in expected value (i.e Large vs Small and Large vs Gamble) B) There was no significant difference in preference for stimuli with the same EV (Small vs Gamble)

3.2.3 Accounting for the effects of visual or motor biases on task performance

Visual biases

In this task, the correspondence of abstract visual stimuli (circle, cross, etc) to different trial conditions was randomised between mice. The purpose of this randomisation was to ensure animals' preferences were driven by learned stimulus-outcome associations rather than by a preference for specific visual stimuli. The four potential abstract visual stimuli used were: a circle, a cross, a horizontal grating, and a vertical grating (shown in Fig 3.4A-B). There was no significant effect of visual stimulus identity on mean accuracy across all 35 mice (2-way ANOVA, $p = 0.27$, Fig 3.4A). Examining the differences in mean accuracy for every combination of stimulus pairs

confirmed mice did not display a significant preference for any given stimulus over another (pairwise t-tests, p = [0.08, 0.13, 0.19, 0.50, 0.92, 0.74], Fig 3.4B). Thus, we found no evidence that visual object identity influenced the animals' preferences, thereby increasing confidence that their choices were guided by the associations between stimuli and outcomes rather than the sensory features of the visual stimuli themselves.

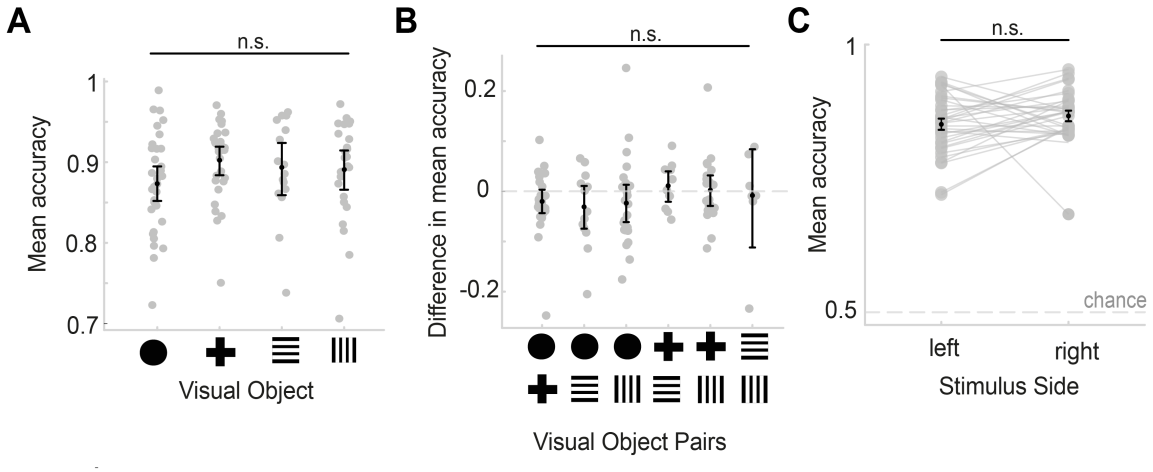


Figure 3.4: Controlling for visual and motor biases. A) Across mice, there was no significant difference in accuracy across the four visual stimuli used. B) Pairwise comparisons of accuracy differences across visual stimuli revealed no significant difference in mean accuracy for any stimulus pair C) There was no significant side bias observed across mice

Motor biases

Mice often displayed significant side biases during the early stages of task acquisition. However, comparing the performance of expert mice on trials where stimuli were presented on the left or the right revealed no significant difference in accuracy overall (paired t-test, p = 0.09, Fig. 3.4C). This suggests that the side biases observed earlier in training did not exert a meaningful influence on expert behaviour.

Taken together, these control analyses suggest that visual and/or motor biases did not have a significant effect on performance in this task.

3.2.4 Task acquisition

A key goal when establishing a novel behavioural paradigm is to ensure the behaviour can be learnt by a large proportion of trained animals and in a reasonable time frame.

To acquire this task, mice underwent a shaped training programme. In the earliest stage of training, they were only presented with one stimulus type (Large) on every trial. The wheel gain - which determines the magnitude of movement required to move the on-screen stimulus - was set to a high level, to facilitate the mice learning to associate the wheel movement with the movement of the on-screen stimulus by only requiring small movements to make a selection. The criterion for progressing to the next stage of training was to correctly select the stimulus on >70% of trials on this simplified version of the task for three days consecutively. Once this initial 'acquisition criterion' was met, the wheel gain was reduced meaning larger and more intentional movements were required to move the stimulus. If the performance of the animal dropped below > 70%, the animal was again required to reach this criterion for three consecutive sessions before additional stimuli were added. Additional stimuli (Small and then Gamble) were added serially when animals' performance was > 70% across all trials for the preceding three sessions (i.e Small was added after three sessions of >70% correct performance on Large trials, and Gamble was added after three sessions of >70% correct performance on Large AND Small trials). The criterion for reaching 'expert-level performance' on this task was to display an overall accuracy of >80% on all trial types. Only mice that reached this 'expert' criterion were used for the behavioural analysis described above, to ensure that differences in accuracy across trial types reflected learned preference rather than inconsistent task performance or ongoing effects of learning.

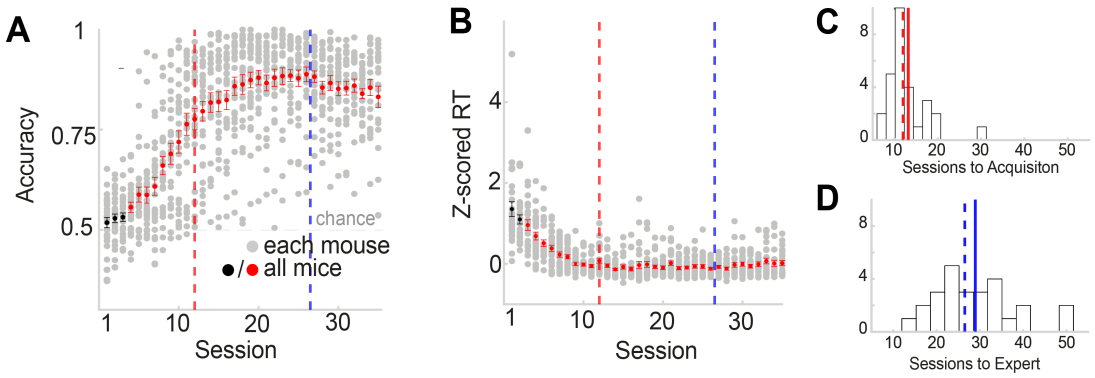


Figure 3.5: Task acquisition A) Overall accuracy across sessions. Grey dots represent individual mice, black/red dots indicate the mean accuracy across mice (error bars, s.e.m). Red dots indicate sessions where performance was significantly different from the first session (pairwise t-test, $p < 0.05$). The red and blue lines indicate the median session number to reach the acquisition (12 sessions) and expert criterion (26.5 sessions) respectively B) Z-scored response times across sessions C) The number of sessions taken to reach the acquisition criterion per mouse (dashed line, median = 12 sessions; solid line, mean = 13.4 sessions) D) The number of sessions taken to reach the expert criterion per mouse (dashed line, median = 26.5 sessions; solid line, mean = 29.3 sessions)

Mice took an average of 12 sessions to reach the initial 'acquisition criterion' (Fig 3.5A, red line and Fig 3.5C). Subsequently, a further 14 sessions were required, on average, to meet the criterion for expert performance (median = 26.5 sessions, Fig 3.5A, blue line and Fig 3.5D). While this duration is longer than that of other, simpler paradigms used to study decision-making in mice, it is comparable to the training time reported in studies that used similarly structured tasks[94;149]. Notably, a common feature across mice was that average response times decreased slightly before choice accuracy increased (paired t-tests, accuracy significantly different from session one ($p < 0.05$) after 4 sessions, (Fig3.5A), reaction time significantly different from session one ($p < 0.05$) - 3 sessions (Fig 3.5B)). This suggests that animals learned to associate wheel movement with reward acquisition before they began to relate movements and outcomes to the on-screen visual stimulus. This is consistent with other work ongoing in the lab investigating individual learning trajectories in mice which has observed a drop in reaction time before a substantial increase in performance[168]. The high proportion of mice that learnt this task (35/38) indicates this is an appropriately high-throughput paradigm, while the consistent rate and learning trajectory mice

followed to acquire this task increases confidence that these behavioural observations are generalisable across mice.

It is tempting, when examining behaviour from the acquisition phase of a task, to consider what these data can reveal about the learning mechanisms involved. For instance, examining changes in performance, response time, and attitude to outcome uncertainty throughout training. However tantalising, this training data does not provide a sound base for answering such questions. The primary reason for this is that the visual stimuli in this task are introduced in series during training. As such, behavioural responses to different stimuli cannot be easily compared during learning. To study learning in this context, the task should be the same from the first day for all mice, similar to recent studies from the lab[168]. This task does not meet these criteria and therefore we must satisfy ourselves with the study of stable, expert behaviour instead.

3.3 Discussion

The central aim of this thesis is to investigate the neural correlates of economic decision-making across multiple regions and projections of the frontal cortex in mice. To achieve this, it was first necessary to establish a behavioural task in which decision variables are systematically manipulated such that their representation in frontal cortical circuits can be studied. Combining this task with techniques for monitoring neural activity thus provides an opportunity to investigate the neural representation of key decision variables during decision-making and goal-directed behaviour in mice.

This chapter describes a novel 2AFC visually-guided decision-making task in which head-fixed mice are trained to make choices in response to abstract visual stimuli, akin to existing tasks in primate literature[12]. Thirty-five mice have been trained to an expert level of performance on this task. Critically, mice displayed significantly higher accuracy for visual stimuli corresponding to a high expected value compared to stimuli associated with a low expected value. These results were consistent across

mice and robust to visual and motor biases. This suggests that mice are sensitive to differences in expected value even when there is only a single stimulus to choose from. In two stimuli trials, where mice were required to make a direct choice between two simultaneously presented stimuli, mice consistently selected the stimulus with the higher expected value. Where the expected value was the same, namely in the Small and Gamble conditions, mice displayed no significant preference.

In summary, the observed differences in behavioural performance depending on reward size or probability are critical for demonstrating that mice are capable of distinguishing between different stimuli and making choices based on the rewards they represent. These results therefore indicate that this task offers a valid means of probing the mechanisms underlying economic decision-making in mice.

3.3.1 Limitations and future directions

Despite its utility in assessing economic decision-making behaviour in mice, the task described above is nonetheless subject to several limitations and possible future refinements.

Firstly, the design of this task is such that mice are only required to respond to one stimulus at a time. In other words, mice are offered a binary choice between a reward-predicting stimulus and nothing. As such, arguably the most straightforward strategy for maximising reward in this task is simply to respond to the presence or absence of a visual stimulus on the screen. This raises a potential concern that this task is capturing reflexive responses driven by coarse sensory processing (i.e stimulus detection), rather than abstract value judgements. However, the observation that task performance and response time scale significantly with expected value suggests that mice are sensitive to the reward contingencies of the presented options. This suggests that this task is eliciting value computations, even when only one stimulus is presented.

The animals used in this study are also highly trained. There are several advantages

to studying highly trained mice, namely that well-learned and stable behaviour minimises variability between sessions and allows patterns of behaviour and, ultimately, neural activity to be compared across different sessions and mice. However, one of the challenges posed by studying over-trained mice is that the learned behaviour is more stereotyped and habitual. Given the frontal cortex is particularly important for flexible, goal-directed behaviour, it may be less engaged in a familiar, well-learned task. As such, studying over-trained mice raises the possibility that subsequent neural recordings may not capture the window in which the frontal cortex is maximally engaged in this task. Nonetheless, similarly to above, the differences in accuracy observed across stimulus conditions indicate the mice are still sensitive to differences in reward magnitude and probability even when highly trained. This task therefore still offers an opportunity to study how expected value and risk are encoded in the frontal cortex, even in over-trained mice.

Another key limitation of this task is that there are an insufficient number of task conditions to construct a complete subjective utility curve for each mouse. With only two possible values of reward magnitude (1.5 µL or 3 µL) and two possible values of reward probability (100% or 50%), it is not possible to comment on mice's subjective valuation of different stimuli or attitude to risk i.e the amount they are willing to 'gamble' for different magnitudes of reward. The small number of task conditions was necessary in this context to ensure an appropriate number of trials per condition for the population decoding pipeline used in Chapter 4. However, future work using this task to probe decision-making behaviour could introduce a larger range of stimuli - and potentially choices between them - to more comprehensively assess subjective preference across mice, as has been done in other studies examining economic decision-making in primates[12;169] and rodents[163;164].

Finally, randomising the mapping between specific visual stimuli and outcomes across mice was necessary to demonstrate that their choices are reliably driven by their preference for stimulus-outcome contingencies, not just salient features of the presented stimuli. However, one potential confound this control cannot account for is the extent to which the mouse visual system can distinguish between these stimuli. While the

differences between the visual stimuli are clear to the human eye, mice have poorer visual acuity[170]. As such, it remains a possibility that certain stimuli appear more 'similar' than others, thereby introducing an element of perceptual uncertainty to their choices. Such a confound could muddy our interpretation of the neural responses to these stimuli, as perceptual uncertainty would introduce an additional dimension to the resulting representations beyond the manipulated task parameters. Indeed, previous work has identified neural correlates of decision confidence in the frontal cortex in response to perceptual uncertainty[171]. A possible future refinement of this behavioural task, therefore, could be to utilise stimuli that vary in parameters known to be particularly salient to the rodent visual system: such as orientation or spatial frequency. Such stimuli have been used in other studies examining mouse decision-making behaviour[149;172]. Despite this concern, the consistent accuracy differences observed across stimulus conditions suggest any perceptual uncertainty or differences in salience did not disrupt animals' ability to distinguish between stimulus conditions in this task.

3.4 Summary

The goal of this chapter was to establish an economic decision-making task in mice that would enable us to investigate the neural correlates of decision-making across frontal cortical regions. We developed a novel 2AFC task that required head-fixed mice to choose between abstract visual stimuli predicting different reward magnitudes and probabilities. The results demonstrate that mice could successfully learn the task, showing high accuracy across task conditions. Critically, accuracy and response times scaled with the expected value of the stimuli - indicating that mice were sensitive to differences in the expected value of presented options. These findings indicate that this task provides a valid means of investigating the neural basis of economic decision-making behaviour in mice.

Chapter 4

Representation of decision-variables across frontal cortical regions during decision-making behaviour

4.1 Introduction

Economic decision-making relies on the ability to infer the value of potential options based on past choices and take actions that maximise reward. Studies have indicated that primates and other mammals are capable of making these complex judgements due to their elaborate frontal cortex. The contribution of the frontal cortex to higher-order cognition has been the subject of scientific interest since the case of Phineas Gage's over 150 years ago. Since then, a raft of lesion[10;11;50;68;173], neuroimaging[53;54;61;174;175], and electrophysiological studies[12;13;16;82;85] have implicated several structures within the frontal cortex in the mechanisms underpinning decision-making and reward-motivated behaviour.

Single neuronal recordings across different sub-regions of the frontal cortex have demonstrated that frontal cortical neurons display highly heterogeneous response profiles. These neurons have been found to signal a diverse range of sensory, cognitive, and behavioural variables relevant to economic decision-making including economic

value[12;17;85;86], expected outcome[105;176], confidence[15;171], upcoming choice[16;94], and reward magnitude and preference[13;75;76;100].

However, previous studies seeking to examine the neural correlates of decision-making in the frontal cortex have typically focused on a handful of neurons in one or a small number of sub-regions of the frontal cortex in isolation. This, combined with the diverse array of behavioural paradigms used to study decision-making behaviour, has made it challenging to directly compare neural representations across frontal sub-regions. Thus, it remains unclear how these diverse representations are structured across the sub-regions of the frontal cortex and how they are integrated to support the computations underlying decision-making behaviour.

This chapter will seek to address this gap by leveraging recent developments in large-scale electrophysiological recording strategies to record from multiple regions of the frontal cortex in mice performing a visually-guided economic decision-making task. By monitoring the activity of a large number of frontal neurons in the same behavioural context, this experiment seeks to provide novel insight into the structure of decision variable encoding across the different sub-regions of the frontal cortex during economic decision-making behaviour.

4.2 Results

4.2.1 Distribution of recorded neurons across frontal cortex

134 Neuropixels recordings were performed across 27 mice trained to an expert level of performance in the visually-guided decision-making task described in the previous chapter. In total, 6,944 high-quality neurons were recorded across these 134 recording sessions. 'High-quality' here refers to well-isolated single units identified following spikesorting (Kilosort[151]) and manual Phy curation displaying a baseline within-trial firing rate > 2 spikes/second and with a presence ratio of 0.9 (meaning the neuron was detectable for at least 90% of the recording).

Region	n recorded	n modulated	n non-modulated	% modulated
ACA	517	462	55	89.4
ILA	769	602	167	78.3
MOs	1521	1380	141	90.7
OFC	1706	1277	429	74.9
PL	1362	1148	214	84.2
AON	628	458	170	72.9
OLF	441	334	107	75.8

Table 4.1: Distribution of recorded and task-modulated neurons across frontal regions

We initially sought to identify the distribution of recorded neurons across different regions of the frontal cortex. This was achieved through probe tract alignment. Coating the Neuropixels probe with DiI before recordings enabled the location of the probe tract to be verified during post-hoc histology(Fig 4.1A). The segments of visible probe tract were used to reconstruct a 3D representation of the probe trajectory (IBL Electrophysiology Atlas), ultimately producing anatomical coordinates for each electrode channel in standard space. The anatomical location of each electrode channel was used to map each recorded unit onto its associated brain region (Fig 4.1B). Table 4.1 outlines the number of recorded neurons found in each region (Fig 4.1C). Overall, the largest proportions of recorded neurons were located in orbitofrontal (1706), secondary motor (1521), and prelimbic cortex (1362).

Next, the proportion of recorded neurons in the frontal cortex that were 'task-responsive' was established. This was achieved by performing a signed rank test comparing trial-averaged firing rate in defined windows around key task events (e.g post-stimulus, choice-start, post-choice etc) to pre-stimulus baseline firing rate (see Appendix A.2 - Identifying task-responsive neurons). Neurons displaying a significant difference in their baseline firing rate and average firing rate in any task window were deemed 'task-responsive'. All time points at and after reward delivery were excluded to ensure task-related activity was not contaminated by reward-driven responses.

Overall, the majority of neurons (5661/6944) across all frontal cortical regions were observed to be task-responsive (82% responsive overall, see Fig 4.1C). Task-modulated activity was not found to be specific to frontal regions, as a large proportion of neurons

in olfactory regions - included as a control - were also observed to be task-responsive. There was no significant effect of brain region on the proportion of task-responsive neurons (one-way ANOVA, p = 1). These relatively non-selective task responses across frontal regions are consistent with recent reports that simple 2AFC tasks drive brain-wide activity changes[155;177].

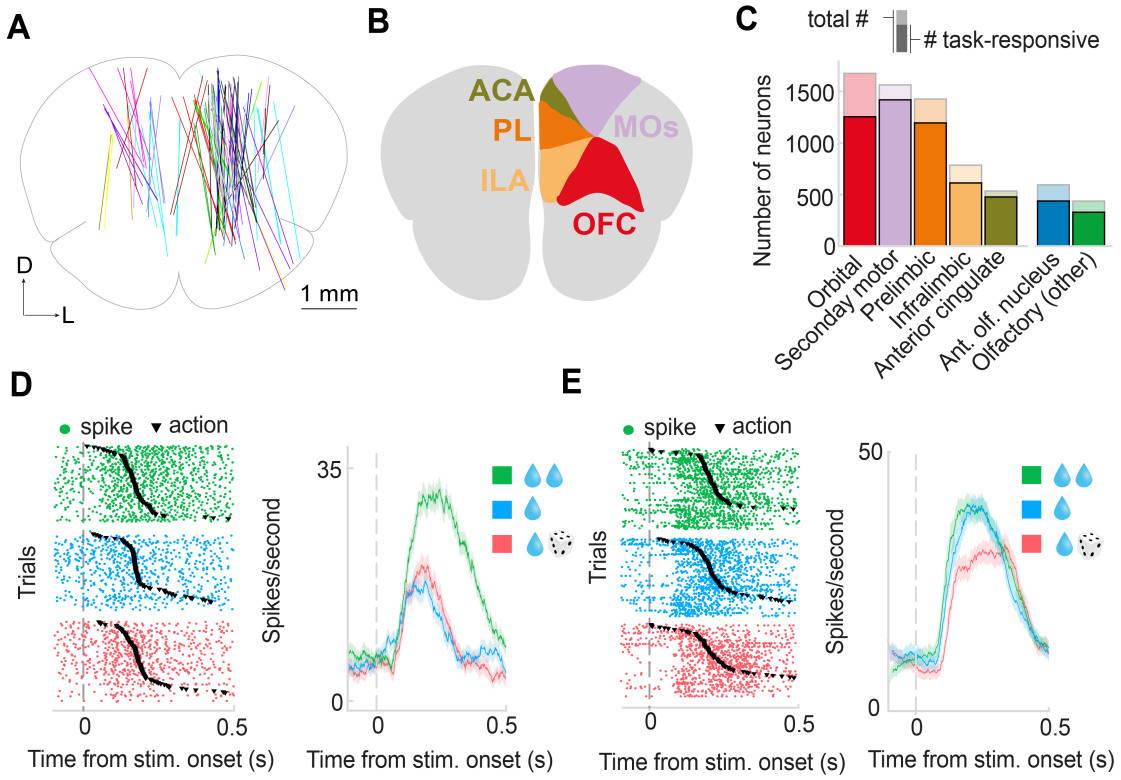


Figure 4.1: Large scale recordings across frontal cortical regions identified individual neurons modulated by task variables A) Distribution of probe tracts for all recordings B) Coronal schematic of frontal regions targeted C) Total number of recorded (light colour) and task-responsive (dark colour) neurons across all recordings. The majority of recorded neurons were task-responsive across all frontal regions (overall % modulated = 82%). There was no significant effect of brain region on the proportion of task-responsive neurons. D-E Raster (L) and PSTH (R) plots of two example neurons. This first neuron (D) displays a greater increase in activity in the response to Large (green) trials, and smaller and similar increases in response to Small (blue) and Gamble (red) trials. This second neuron (E) displays large and comparable increases in activity in response to Large (green) and Small (blue) trials, and a smaller increase in response to Gamble (red) trials.

4.2.2 Individual neurons are differentially modulated by task conditions at stimulus onset

Visualising trial-averaged responses for individual neurons revealed a subset of neurons differentially modulated by task conditions. Consider the two example neurons in Fig 4.1D-E. Both neurons appear time-locked to stimulus presentations, showing a consistent and trial non-specific increase in firing rate with a fixed delay after stimulus onset that does not vary with response time. These neurons also display differential responses to different trial conditions. The first neuron displays a robust increase in firing rate in response to Large trials and smaller and comparable increases in firing rate in response to Small and Gamble trials. The former stimulus has a high expected value, the latter stimuli have lower and identical expected valued. Conversely, the second example neuron displays a robust and comparable increase in firing rate in response to the Large and Small trials - both guaranteed rewards - and a smaller increase in firing rate in response to the Gamble stimulus. This indicates a subset of recorded neurons are differentially responsive to different task conditions, perhaps reflecting differences in expected value and risk across stimuli. This is consistent with previous literature, which has identified signatures of expected value and reward probability in individual neurons across frontal regions[12;85;164;178].

4.2.3 Establishing the appropriate time window for investigating stimulus responses

The previous section identified a subset of individual frontal neurons whose activity appeared to correlate with the two key variables manipulated in this task: expected value and risk. To quantitatively demonstrate that neurons in the frontal cortex encode these variables, and to take advantage of the large number of recorded neurons in this dataset, we implemented linear population decoding to assess the extent to which task conditions could be decoded from the activity of recorded frontal neurons (see General methods and materials, Section 2.7.1 - Population decoding).

To dissociate changes in neural activity driven by differences in task conditions, rather than non-specific stimulus- and action-related signals, a kernel regression model (Fig 4.2C) with a single stimulus kernel and a single action kernel was fit to trial-by-trial responses of each neuron (see Steinmetz et al, 2019[155]). Subtracting cross-validated model predictions from the raw single neuron data thus removed non-specific stimulus and action-related activity while retaining condition-specific activity. All subsequent decoding analysis was performed on the residual spiking activity (Fig 4.2D).

We first sought to determine a suitable time window relative to stimulus onset in which trial conditions (Large, Small, Gamble) could be reliably decoded from the pooled activity of all task-responsive neurons. To achieve this, a 10-fold cross-validated three-class Support Vector Machine (SVM) was trained and tested to classify these trial conditions in 10 ms time bins from 50ms before to 350ms after stimulus onset. As noted above, all time points at or following reward delivery were excluded to ensure differences in activity across task conditions were not driven by differences in the resulting reward (i.e the volume of water). The resulting temporal generalisation matrix (Fig 4.2A) revealed a significant cluster of strong decoding in a 150ms to 350ms window after stimulus onset. This suggests that trial conditions could be successfully decoded from neural activity in this window after stimulus onset. Thus, this is a suitable window to examine how population activity in frontal neurons encodes differences in task conditions.

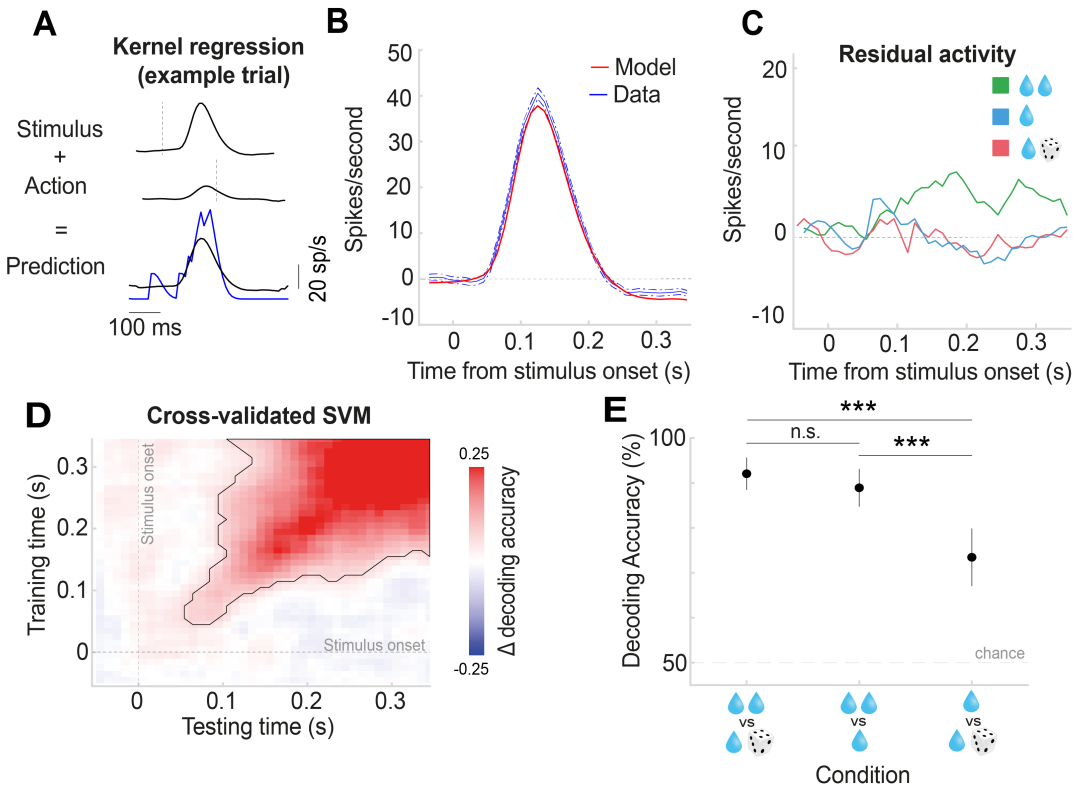


Figure 4.2: Population activity at stimulus onset reflects expected value A) An example stimulus and action kernel being used to predict an example neuron's spiking activity. Bottom panel - an example fit of spiking data using these two different kernels during one trial (stimulus and action). The blue trace denotes the neuron's raw spiking data smoothed with a causal filter and the black line represents the prediction of the model. B) Average empirical stimulus-locked response (blue) and model prediction (red) for the same neuron C) Residual activity after subtracting model predictions from the raw spiking data. The residuals capture differences in spiking activity across task conditions D) A three-class SVM was used to identify a time window in which trial conditions could be reliably decoded from the pooled activity of all task-responsive neurons. This temporal generalisation matrix comparing decoding accuracy in 10ms bins for testing and training data reveals a window of strong decoding 150ms to 350ms post-stimulus-onset. E) Binary decoding of trial conditions using pooled population activity between 150 and 350ms post-stimulus onset. Decoding accuracy was significantly higher when classifying trials with different expected values (Large vs Gamble and Large vs Small) compared to the same expected value (Small vs Gamble)

4.2.4 Population activity across the frontal cortex reflects differences in expected value

Having established an appropriate time window for studying stimulus-driven responses, binary classifiers were then implemented to investigate which pairs of trial conditions contributed to the strong decoding observed in this time window.

Overall, a high level of decoding accuracy was observed across stimulus conditions (Fig 4.2E). All three binary comparisons - Large vs Gamble, Large vs Small, and Gamble vs Small - were decodable significantly above chance (permutation test, p values < 0.00001). Notably, greater decoding accuracy was observed when comparing trial conditions with different expected values (Large vs Gamble - 92.1%; Large vs Small - 88.9%) compared to trial conditions with the same expected value (Small vs Gamble, 73.4%). The difference in decoding accuracy between Large vs Small and Gamble vs Small and between Large vs Gamble and Gamble vs Small was significant (permutation test, p-values < 0.00001). This indicates that decoding performance is significantly better when comparing conditions with different expected values compared to conditions with the same expected value. The difference in decoding accuracy between Large vs Small and Large vs Gamble was not significant (permutation test, p-value $= 0.0740$). These differences in decoding accuracy cannot be explained by differences in the visual features of the presented stimuli, as the mapping between visual stimuli (Cross, Circle etc) and reward contingencies (Large, Small, Gamble) was randomised between mice. Thus, population activity across the frontal cortex reflects differences in task conditions, particularly those corresponding to different expected values, with a high level of accuracy.

4.2.5 Expected value and risk encoding in frontal cortical regions

The behavioural task used in this experiment presents mice with abstract visual stimuli that vary in their expected value and risk. As such, binary comparisons between the Large and Small conditions, which have different expected values but

the same level of risk, and the Small and Gamble conditions, which have the same expected value and different levels of risk, offer an opportunity to test how expected value and risk are encoded in frontal cortical neurons. Existing literature has found evidence that neurons in various regions of the frontal cortex encode expected value and risk, alongside other economic variables, using a variety of behavioural tasks[12;17;81;83;85;86;90;103;104;115;129;165;176;179]. Thus, a key question that emerges here is: how does the encoding of expected value and risk compare across the subregions of the frontal cortex in this behavioural context?

To address this, binary classifiers comparing neural activity across pairs of task conditions (Large vs Small and Small vs Gamble) were trained and tested separately for each population of region-defined neurons. One challenge with this analysis is that certain frontal regions were over-represented in the dataset, largely due to the trajectory of these recordings and the relative size of different frontal cortical structures. For instance, 24.4% of recorded neurons were from orbitofrontal cortex compared to just 7.4% of neurons in infralimbic. To account for this, decoding was performed on matched samples of neurons (50 - 800 neurons, in 50 neuron bins) to both compare decoding accuracy across brain regions and determine how decoding accuracy changes with increasing neuronal sample size.

Decoding accuracy for both EV (i.e Large vs Small) and risk (i.e Small v Gamble) comparisons was significantly above chance (permutation test, $p < 0.05$) across all regions tested (Fig 4.3A-B). The relative encoding across regions for both comparisons was also consistent across neuronal sample sizes. In the case of expected value, anterior cingulate cortex (ACA) displayed the highest decoding accuracy - reaching a maximum decoding accuracy of 81.9% accuracy for 350 neurons. At a matched number of neurons, strong decoding accuracy was also observed in MOs (71.95%, max. =78.5% for 550 neurons), PL (70.17%, max. =78.8% for 800 neurons), and OFC (68.5%, max. = 74.9% for 750 neurons). While decoding accuracy was generally lower in the case of risk - consistent with the results of population decoding outlined in Figure 4.2B - all tested regions nonetheless performed significantly above chance (permutation test, $p < 0.05$). The highest decoding accuracy for risk was observed in

ACA and PL, reaching a maximum of 65.7% and 65.4% for 350 neurons respectively (Pl max. = 69.4% for 800 neurons). Strong decoding accuracy was also observed in ILA (60.0%, max. = 62.2% for 450 neurons), and OFC (58.7%%, max. = 63.2% for 750 neurons). Notably, the relative encoding across different regions was different between EV and risk comparisons. For example, MOs displayed the second-highest decoding accuracy in the EV comparison (max. = 74.9% for 550 neurons) but the lowest decoding accuracy in the risk comparison (max. = 55.7% for 550 neurons).

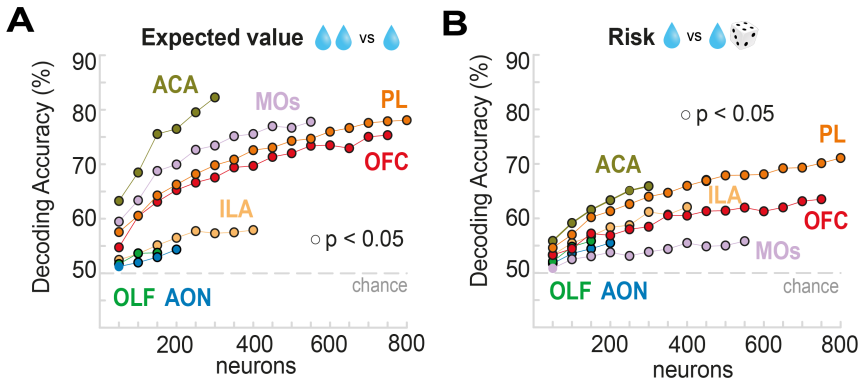


Figure 4.3: Binary decoding of expected value and risk across frontal cortical regions A) Average decoding accuracy for the EV binary classifier (i.e Large vs Small). The binary classifier was implemented on populations of region-defined neurons (50 to 800 neurons) to compare EV coding across frontal subregions. Data points with a black outline are significantly above chance (permutation test, $p < 0.05$) B) Average decoding accuracy for the risk binary classifier (Small vs Gamble). Data points with black outline are significantly above chance (permutation test $p < 0.05$)

The results of the binary decoding output reveal several interesting observations. Firstly, it appears that stimulus encoding is consistently strongest - for both EV and risk - in the dorsomedial regions of frontal cortex, namely ACA and PL. Other regions displayed differential encoding of these two variables. For example, MOs appeared to strongly represent differences in EV, but not risk, while ILA displayed stronger decoding accuracy for risk but not EV. The lowest decoding accuracy was observed in the olfactory regions (OLF and AON), indicating these regions do not strongly encode either variable. Overall, this observation suggests that different sub-regions of the frontal cortex do indeed differentially represent specific features of decision-making. However, it challenges the suggestion that certain regions are highly specialised in the variables they represent: as even frontal cortical regions with 'lower' EV coding

displayed decoding accuracy significantly above chance e.g infralimbic. Interestingly, the results of this binary decoding also allude to a spatial distinction between expected value and risk representation in the frontal cortex. EV appeared, on the whole, to be better represented in the dorsomedial and dorsolateral regions of the frontal cortex (ACA, PL, MOs, OFC), while risk appeared better represented in medial regions (ACA, PL, ILA) of the frontal cortex.

4.2.6 Spatial distribution of variable encoding in frontal cortex

The results of the region-specific decoding point to a possible spatial structure underlying the encoding of expected value and risk across the sub-regions of the frontal cortex. In particular, dorsal regions (ACA, MOs, PL) displayed the highest decoding accuracy for the EV comparison, while medial regions (ACA, PL, ILA) displayed the highest decoding accuracy for the risk comparison. To further investigate the spatial structure of this variable encoding, we next sought to examine how decoding accuracy changed along the mediolateral (ML) and dorsoventral (DV) axes of the frontal cortex.

The binary decoding implemented in previous sections assigns each neuron included in the analysis an activation weight. The sign of the weight indicates the neuron's relative selectivity for the two conditions being compared, while the magnitude indicates how strongly the activity of that neuron can be used to classify them. To examine how population encoding of EV and risk varied across the frontal cortex, absolute activation weights were visualised as a function of the neurons' spatial location, rather than region-label, yielding 2-dimensional spatial activation maps (see General methods and materials, Section 2.7.1 - Spatial binary decoding). These maps provide a visual representation of the spatial distribution of neural activation, highlighting areas with relatively higher (red) and lower (blue) average activation weights. This approach allows us to investigate the spatial organisation of neural representations of EV and risk beyond brain region labels, focusing on their distribution along the

mediolateral (ML) and dorsoventral (DV) axes of the frontal cortex.

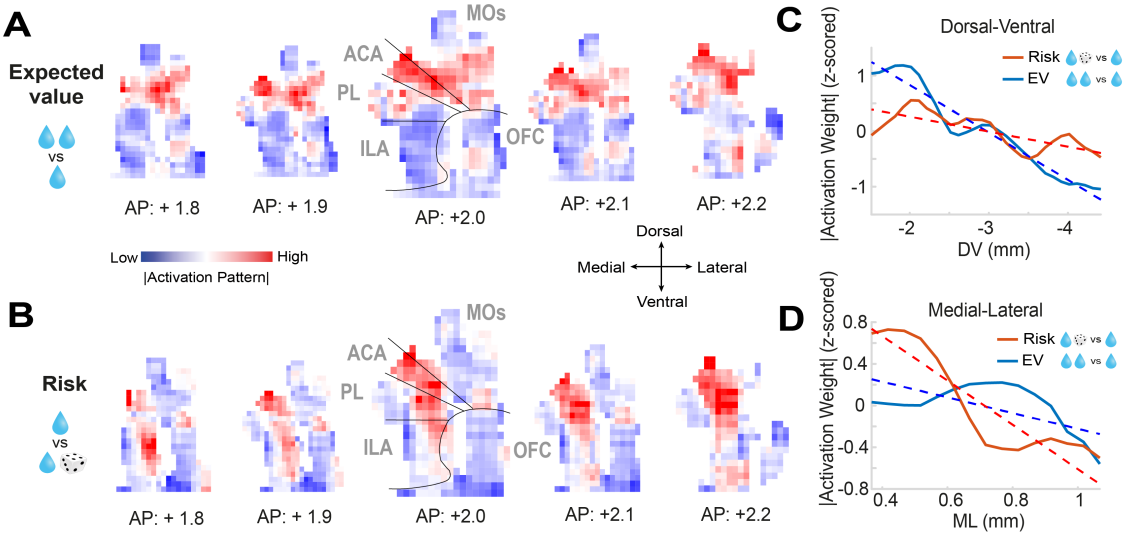


Figure 4.4: Distinct spatial gradients for EV and risk coding across frontal cortex A) Spatial decoding plot showing absolute EV classifier weights averaged in spatial bins with an atlas image overlay. Distinct dorsal-to-ventral gradient for expected value representations was observed and maintained across the anterior-to-posterior axes (left-to-right) B) Spatial decoding plot showing absolute risk classifier weights averaged in spatial bins with an atlas image overlay. Distinct medial-to-lateral gradient for risk representations was observed and maintained across the anterior-to-posterior axes (left-to-right) C) Activation weights (z-scored) across the dorsal-ventral axes split by risk and EV coding. The dotted line represents the slope of the change in the activation weights. D) Activation weights (z-scored) across the medial-ventral axes split by risk and EV coding

Strikingly, these maps revealed distinct regions of activation for EV and risk (Fig 4.4A-B). For EV coding, high activation was observed in dorsal regions of frontal cortex - in both dorsomedial and dorsolateral areas - but not in ventral regions. Conversely, risk coding was observed most strongly in medial regions - both dorsal and ventral - but not in lateral regions. Moreover, these spatial patterns were consistently observed irrespective of the limited range of anterior-posterior coordinates recorded in this study.

To assess whether these observed differences in activation for EV and risk were statistically significant, activation weights were averaged in the medial-to-lateral and dorsal-to-ventral dimensions and compared between variables. Averaging activation weights across DV coordinates (Fig 4.4C) revealed that EV displayed a statistically

significant reduction in activation in the dorsal-to-ventral direction (permutation test, p-value < 0.00001), while risk did not (permutation test, p-value $= 0.086$). In contrast, averaging across ML coordinates (Fig 4.4D) revealed that risk coding displayed a significant reduction in activation in the medial-to-lateral direction (permutation test, $p = 0.012$), while EV did not (permutation test, p-value $= 0.69$). The change in activation was found to be significantly different between EV and risk in both dimensions (dorsal to ventral difference - permutation test, $p < 0.00001$; medial to lateral difference - permutation test, $p < 0.016$).

Overall, we observed distinct and significant spatial gradients underlying the encoding of expected value and risk across the frontal cortex. Expected value coding displayed a dorsoventral gradient, with strong coding in dorsal but not ventral regions of the frontal cortex, while risk coding displayed a mediolateral gradient, with strong coding in medial but not lateral regions of the frontal cortex. These results reveal a novel spatial structure underlying decision variable coding in the frontal cortex.

4.2.7 Lateralised neural responses

In this task, visual stimuli are presented on either the left- or right-hand side of the screen on each trial. Thus, in addition to the manipulations in expected value and risk, stimulus location is also manipulated on a trial-by-trial basis.

Examining single-neuron responses to stimuli presented on the left and right revealed a subset of neurons exhibiting highly lateralised responses to contralateral and ipsilateral stimulus presentation (relative to the recorded hemisphere). For instance, Figure 4.5A-B illustrates a neuron more strongly modulated by contralateral stimuli than ipsilateral stimuli. Functional and representational lateralisation have been observed in various brain regions and circuits[180;181]. The lateralised responses observed across a subset of recorded neurons in this study raised certain questions. Firstly, how strongly is stimulus location encoded in the activity of frontal cortical neurons? And, secondly, to what extent does the location of a stimulus influence or interact with the encoding of EV and risk described above?

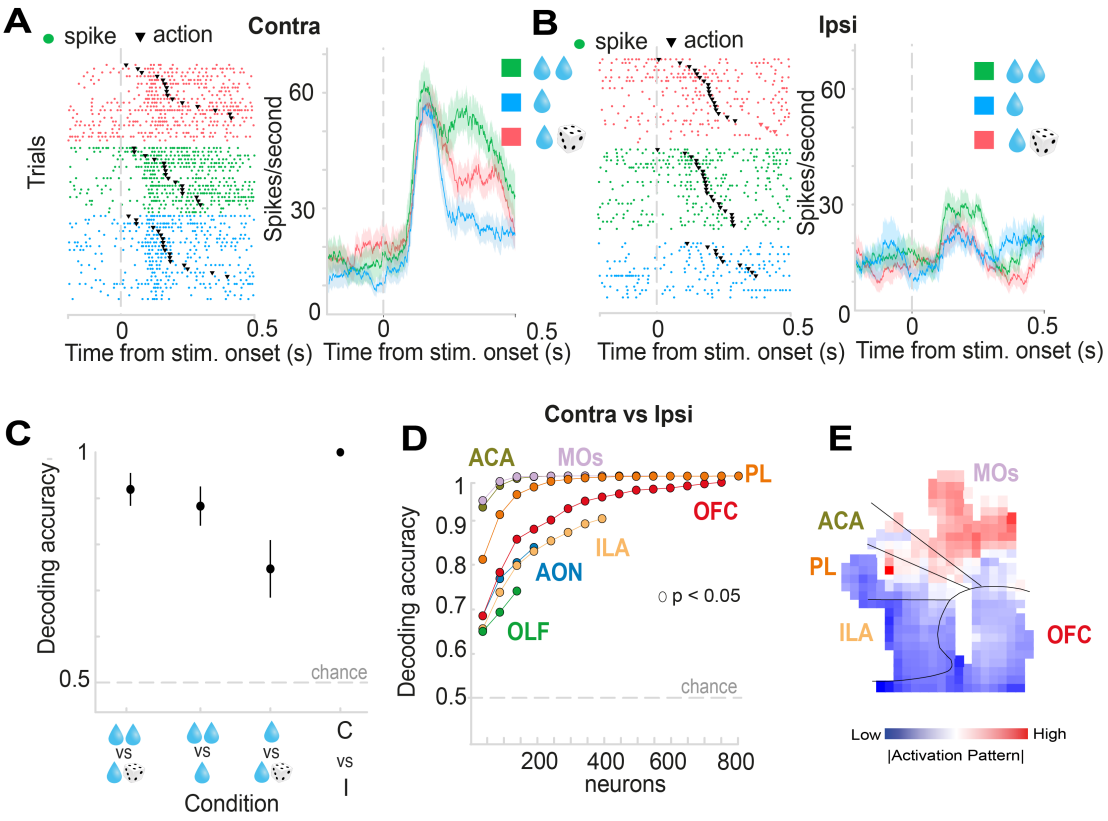


Figure 4.5: A) An example neuron's response to contralateral stimulus presentation, raster (left) and PSTH (right) B) The same example neuron's response to ipsilateral stimulus presentation, raster (left) and PSTH (right) C) Binary decoding of trial conditions (Large v Gamble, Large v Small, Gamble v Small) and contralateral vs ipsilateral stimulus location based on pooled population activity. Contra vs ipsi decoding displayed 100% decoding accuracy D) Per region decoding of stimulus location (contra vs ipsi). High decoding accuracy was observed across all recorded regions, particularly ACA, MOs, and PL E) Z-scored absolute activation weights across a 2D map of frontal cortex revealed strong stimulus location coding, particularly in dorsal regions

A binary classifier was implemented to assess the extent to which stimulus location (ipsilateral vs contralateral to the recorded hemisphere) could be decoded from frontal cortical population activity in the same time window as the analysis above (150ms to 350ms post-stimulus). Strikingly, the decoding accuracy for differentiating ipsilateral versus contralateral stimuli reached 100% (Fig 4.5C), higher than any of the binary comparisons between trial conditions. This strong coding of stimulus location suggests a significant degree of lateralisation in task-modulated population activity across the frontal cortex.

Examining region-specific coding of stimulus location revealed all tested regions encoded stimulus location significantly above chance (permutation test, $p < 0.05$). This encoding was largely driven by MOs (100% for 350 neurons), ACA (100% for 350 neurons), and PL (99.0% for 350 neurons, max. = 100% for 650 neurons), although all recorded regions reached a high level of decoding accuracy (Fig 4.5D). The pronounced encoding in MOs suggests that the representation of location may be related to motor planning for upcoming responses. Finally, visualising stimulus location encoding across ML and DV coordinates by plotting 2D spatial maps of absolute activation weights, as above, demonstrated that dorsal regions of cortex - both medial and lateral - most strongly encoded stimulus location (Fig 4.5E).

These results indicate that population activity in the frontal cortex strongly encodes the location of presented stimuli. Strong encoding of stimulus location was observed across all regions of the frontal cortex, however it was particularly pronounced in dorsal regions such as MOs and ACA. Together these observations support strong encoding of stimulus location in dorsal cortex during economic decision-making.

4.2.8 Interaction of stimulus location and economic variable encoding

Having identified a strong representation of stimulus location, particularly in dorsal regions of the frontal cortex, we next sought to investigate how this encoding interacts with the observed encoding of EV and risk.

Performing binary decoding of stimulus identity separately for ipsilateral and contralateral trials revealed greater decoding for contralateral trials across all comparisons (Fig 4.6A-B). For example, decoding accuracy for the 'Large vs Gamble' condition was 98.4% (permutation test, p-value < 0.0001) for contralateral trials, compared to 73.8% (permutation test - p-value = 0.01) for ipsilateral trials. Similarly, decoding accuracy for the 'Large vs Small' condition was 94.9% (permutation test, p-value $<$ 0.0001) for contralateral trials, and 69.2% (permutation test - p-value = 0.029) for ipsilateral trials. Finally, decoding for the 'Small vs Gamble' condition was 82.3%

(permutation test, p-value < 0.0001) for contralateral trials, but not significantly different from chance for ipsilateral trials (48.7%, permutation test, p-value $= 0.446$). Notably, it remained the case that stimulus comparisons with different expected values displayed greater decoding accuracy than stimulus comparisons with the same expected value for both ipsilateral and contralateral trials. Thus, while the relative decoding accuracy across binary comparisons was conserved, decoding accuracy was found to be much higher across all conditions for contralateral trials.

We next sought to determine how this contralateral bias affects the spatial structure of expected value and risk coding across the frontal cortex. To achieve this, binary classifiers were used to generate activation weights for EV and risk coding separately for contralateral and ipsilateral trials. Z-scored absolute activation weights were then visualised on the same spatial maps as before (Fig 4.6C-D). For contralateral EV and risk coding (Fig 4.6C), we observed the same - and indeed stronger - activation patterns across the frontal cortex as was observed when the decoding was run on all trials (ipsi and contra trials combined). EV coding still displayed a reduction in coding in the dorsal-to-ventral direction, and risk a reduction in the medial-to-lateral direction. This pattern of activation was not observed for ipsilateral trials (Fig 4.6D). While some EV coding was also identifiable in the dorsal regions for ipsilateral trials, albeit much weaker than for contralateral trials, there was no clear pattern of activation for risk coding in ipsilateral trials. This is consistent with the observation that the 'Small vs Gamble' binary decoding accuracy for ipsilateral trials was not significantly above chance.

Testing the statistical significance of these coding gradients along the DV and Ml axes confirmed that the patterns of activation for EV and risk for contralateral trials followed the same distribution as the overall activation patterns for all trials. The EV activation gradient was found to be significant in the dorsoventral direction (permutation test, p-value < 0.00001), the risk activation gradient significant in the mediolateral direction (permutation test, p-value < 0.00001), and both pairs of gradients were significantly different from each other (difference-ML, permutation test, p-value < 0.00001; difference-DV, permutation test, p-value $= 0.0080$). For ipsilateral trials,

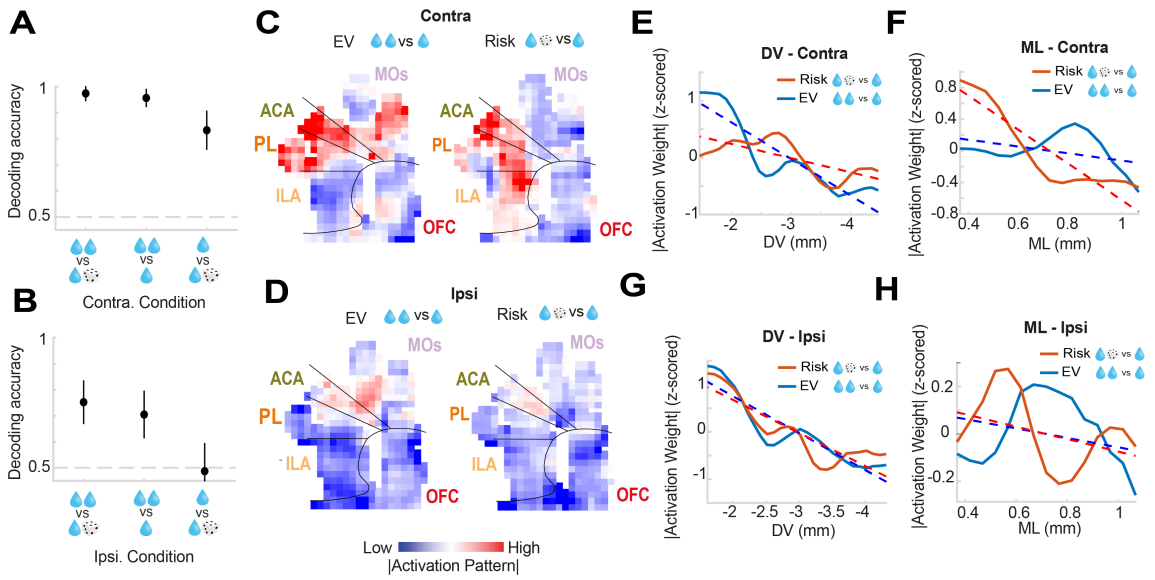


Figure 4.6: Ipsilateral and contralateral coding of EV and risk A) Population binary decoding of trial condition pairs for contralateral trials only B) Population binary decoding of trial condition pairs for ipsilateral trials only C) 2D maps of z-scored absolute activation weights for EV (left) and risk (right) coding - contralateral trials only. Contralateral coding displays the same spatial structure as all trials (ipsi + contra) combined C) 2D maps of z-scored absolute activation weights for EV (left) and risk (right) coding - ipsilateral trials only. Ipsilateral coding was much weaker than contralateral coding. E-F Gradients for change in activation weights across DV (E) and ML (F) coordinates for contralateral trials. EV displayed a significant DV gradient and risk displayed a significant ML gradient, consistent with the gradients observed for all trials combined. G-H) Gradients for change in activation weights across DV (G) and ML (H) coordinates for ipsilateral trials. None of these gradients were significant or significantly different.

the only significant gradients were observed in the dorsoventral direction for both EV and risk (EV, permutation test, p-value < 0.00001; risk, permutation test, p-value < 0.00001). However, these were not found to be significantly different from each other (difference-DV, permutation test, p-value $= 0.19$), indicating this was a general reduction in coding rather than a specific change in either variable. None of the ipsilateral activation gradients in the medial-to-lateral direction were found to be significant for either variable (EV-ML p-value $= 0.83$, risk-ML p-value $= 0.73$). This indicates that the dorsal-to-ventral gradient for EV coding and medial-to-lateral gradient for risk coding observed across the frontal cortex are almost exclusively driven by trials in which stimuli were presented contralaterally to the recorded neural population.

Taken together, these results suggest that representations of decision variables - namely EV and risk - across the frontal cortex in mice are highly lateralised.

4.3 Discussion

The goal of this chapter was to examine the representation of decision variables across different regions of the frontal cortex during decision-making behaviour in mice. Previous work has identified neural correlates of a variety of sensory, cognitive, and behavioural variables across different sub-regions of the frontal cortex in the context of decision-making[12–17]. However, these studies typically focused on one or a small number of brain regions in isolation using a disparate array of behavioural tasks. This experiment sought to provide novel insight into the structure of decision variable representation in the frontal cortex by recording the activity of a large number of neurons across different regions of the frontal cortex in the same behavioural context.

Neuropixels probes were used to record the activity of almost 7,000 frontal cortical neurons in mice performing a visually-guided decision-making task. The majority of neurons were found to be 'task-responsive' across all recorded regions. Implementing linear decoding on population activity following stimulus-onset - the task epoch in which information about the presented stimuli must be integrated to drive an appropriate response - revealed that different trial conditions could be decoded with a high degree of accuracy, particularly when comparing conditions that had a different expected value (i.e 'Large vs Small' and 'Large vs Gamble'). This indicates that population activity in the frontal cortex reflects differences in expected value.

Examining the extent to which different frontal areas contributed to this decoding revealed a dissociation in the regions encoding the two key variables manipulated in this task: expected value (EV) and risk. While some regions of the frontal cortex displayed strong encoding of both EV and risk, most notably anterior cingulate cortex (ACA) and prelimbic cortex (PL), other areas were found to encode one or the other variable more strongly. For example, secondary motor cortex (MOs) was found

to strongly represent EV but not risk, while infralimbic cortex (ILA) was found to strongly encode risk but not EV. The differential coding of EV and risk observed across regions also pointed to a potential spatial structure underlying the representations of these variables: as medial regions of the frontal cortex (ACA, PL, ILA etc) appeared to display stronger encoding of risk, while dorsal regions appeared to more strongly encode EV (ACA, MOs, PL, OFC).

The rich representation of EV, risk, and stimulus location observed in ACA is consistent with previous work identifying ACA as a critical node for integrating multiple decision variables during goal-directed behaviour[69;82;87;90;92;182]. These observations give further credence to the theory that ACA contributes to the generation of a comprehensive representation of task space to support high-level and abstract control over decisions and actions.

One notable discrepancy between the decoding results presented here and previous studies investigating the neural basis of economic decision-making is the relatively modest encoding observed in the orbitofrontal cortex (OFC). As discussed in the Introduction (see 1.2 Single neurons in frontal cortex encode economic variables), OFC is often highlighted as a critical structure underpinning economic decision-making in humans, primates, and rodents[10;12;115]. Therefore, it is somewhat surprising that we observed weaker encoding of EV and risk in OFC compared to other regions of the frontal cortex. This finding suggests that other frontal cortex regions, particularly dorsomedial regions such as the anterior cingulate area (ACA) and prelimbic cortex (PL), may play a more significant role in representing certain decision variables than previously recognised. Nonetheless, it is important to note that even regions displaying 'weaker' encoding of EV or risk in this experiment, including the OFC, still performed significantly above chance levels. Hence, while OFC encoding was relatively less pronounced compared to other structures, it still displayed significant encoding of both variables.

Furthermore, there is evidence that OFC may be more critical for the acquisition or flexible updating of goal-directed behaviour than the maintenance of well-learned

behaviours. For example, Wang et al(2020)[179] used 2-photon imaging to study the activity of OFC and mPFC in mice learning an olfactory discrimination task. They observed that the representation of presented stimuli in OFC diminished throughout training, while the representations that emerged in mPFC remained persistent throughout. The authors suggest that OFC and mPFC may be sequentially recruited to support the acquisition of appetitive associations. Thus, it is also possible that the modest representation of EV and risk observed in OFC, and the relatively strong representation in dorsomedial frontal regions such as ACA and PL, could be explained by our use of highly-trained mice. If OFC is more involved in the acquisition of goal-directed behaviours, reflecting its established role in the formation of stimulus-outcome associations, we may be missing the phase in which OFC is most engaged in this task due to studying neural activity in highly trained mice. Follow-up work examining the encoding of EV and risk in OFC in a comparable task, but focusing specifically on the acquisition phase, will be required to address this alternative theory.

Having identified a difference in the frontal regions most strongly encoding EV and risk, we next sought to investigate the spatial structure of these representations by generating 2D spatial maps of EV and risk coding across the ML and DV axes of the frontal cortex. Examining the patterns of activation across these maps revealed striking and distinct spatial gradients for EV and risk coding. EV coding was found to display a strong dorsoventral gradient, with stronger encoding observed in dorsal than ventral regions, while risk was found to display a strong mediolateral gradient, with stronger encoding observed in medial than lateral regions. These observations are consistent with the results of the region-specific decoding, which observed strong encoding of both variables in dorsomedial structures (ACA and PL) but divergent patterns of encoding elsewhere. These gradients were found to be statistically significant, held across different AP coordinates, and were almost exclusively driven by responses to contralateral trials relative to the recorded hemisphere. The highly lateralised nature of variable encoding observed in this study suggests that mice do not follow a 'goods-based' model of economic choice, as has been identified in primates[84].

This model of economic decision-making suggests that neurons are tuned to the abstract subjective value of different options, rather than to the specific attributes or sensory features of the goods. Our results indicate that, unlike primates, mice encode economic variables in a spatial frame of reference - perhaps reflecting the formation of strong associations between reward-predictive stimuli and the actions required to secure those rewards.

The identification of distinct spatial gradients for EV and risk coding was striking but nonetheless consistent with existing observations of variable representation in the frontal cortex. As noted above, all regions recorded in this experiment were found to encode EV and risk significantly above chance - consistent with previous work noting that neurons responsive to almost all task-variables have been identified across the structures that make up frontal cortex[85;98]. However, the existence of these gradients also provides evidence for representational and perhaps functional differences across different frontal regions. Thus, the results of this study offer an opportunity to reconcile existing observations of frontal cortical function at a single-cell and a population level: single neurons do indeed represent a wide array of variables across the frontal cortex, but it is still possible to resolve differences between regions when you examine how these representations are structured across the whole population. In essence, sub-regions of the frontal cortex do differentially represent fundamental decision variables, but these are differences of degree, not of kind.

4.3.1 Limitations and future directions

The results of this study offer an obvious and eminently testable hypothesis for future work examining the neural basis of economic decision-making. The existence of a dorsal-to-ventral gradient for EV coding and medial-to-lateral gradient for risk coding suggests that disrupting dorsal cortical activity may disproportionately affect value-based decision-making, while disrupting medial frontal cortical activity is more likely to affect risk-based decision-making. We hypothesise that these effects are likely to be highly lateralised (i.e affecting mostly contralateral choices), given the strength of

contralateral EV and risk coding relative to ipsilateral coding. Unfortunately, this follow-up experiment would vastly exceed the time frame available for this project. Nonetheless, this represents an exciting opportunity for follow-up work seeking to test how the spatial structure of decision variable representation in the frontal cortex relates to observable behaviour.

One limitation of the approach used in this chapter is that the formation of 2D spatial activation maps based on acute recordings across a large number of mice makes it challenging to determine how the spatial structure of variable encoding compares between mice. It would be impossible to populate a complete activation map for each animal, due to the limited number of acute recordings it is possible to perform in a single craniotomy while maintaining good tissue quality. As such, the results above assume the gradients observed when neural activity is pooled across sessions arise from a comparable coding structure across mice. It remains a possibility, therefore, that the gradients we observe are driven by a subset of mice - or, in the most extreme case, an artefact of pooling across mice - rather than being an inherent feature of economic variable representation in the frontal cortex. However, analysis conducted by a collaborator has demonstrated that neurons recorded simultaneously across different sub-regions display comparable differences in EV and risk coding (not shown) to those observed in the region-specific decoding across the whole population of neurons. That neurons recorded simultaneously in specific regions follow the same trends as the whole population suggests that we would observe the same distribution of EV and risk coding if it were possible to generate unique coding maps per mouse. However, until it is possible to image large 2D or 3D sections of cortex in vivo in the same mouse with high spatial and temporal resolution, reaching a definitive conclusion on how the spatial structure of variable encoding across the frontal cortex compares between mice is unfortunately out of reach.

In addition, the analysis described in this chapter focuses on variable encoding in a 150 to 350ms window post-stimulus onset, the window in which trial conditions are maximally decodable from pooled population activity. However, the time scale of this task makes it difficult to disentangle stimulus and action-modulated neural responses.

There is only a short delay of 0.2 - 0.3 seconds between the stimulus appearing on the screen and the mouse being able to move the stimulus to enact its choice. As a result, while the neural data described above reflects activity at the time of 'stimulus-onset', it is unclear to what extent these signals reflect visually-guided, stimulus-driven responses or choice-related responses. We do not believe this affects our core findings, however, as the distinct gradients observed for EV and risk coding cannot be explained by overlapping stimulus and choice signals. Nonetheless, adapting the structure of the task to incorporate a longer post-stimulus quiescence period in which no movement is permitted may be a useful refinement for future work seeking to compare neural activity in these two task epochs. Furthermore, given that the analysis in this chapter only focuses on pre-reward responses on correct trials, future analyses could also seek to explore the structure of variable representation during other features of decision-making in this task, such as post-reward responses, error trials (though they are few in this well-learned paradigm), or different behavioural states e.g engaged versus disengaged.

Finally, a linear decoding approach was implemented to provide an easily interpretable read-out of neural encoding of key task variables in this experiment. For this kind of variable decoding to succeed the information must be structured in the brain in a manner that the decoder can exploit[183]. As such, a limitation of this approach is that it might fail to detect information encoded in a more complex manner: such as integrated representations of multiple variables. We attempted to address this question by examining the correlation in classifier weights for each manipulated variable (EV, risk, or stimulus location). We did not observe a strong overlap of variable encoding for any of these comparisons (Appendix A.3, Correlating variable encoding). However, an alternative approach to studying high-dimensional neural activity data could be to use dimensionality reduction and unsupervised learning approaches such as spectral clustering. Such an approach might provide better insight into the parallel processing of variables in overlapping populations of neurons. This therefore represents a goal for future work.

4.3.2 Summary

This chapter outlines an experiment in which Neuropixels probes were used to record the activity of thousands of neurons in the frontal cortex of mice performing a visually-guided economic decision-making task. Linear decoding identified distinct spatial gradients for the encoding of risk and EV, largely driven by contralateral trials. These data reveal a novel spatial structure to decision variable representation across the frontal cortex.

Chapter 5

Representation of decision-variables in projection-defined frontal cortical neurons during decision-making behaviour

5.1 Introduction

The frontal cortex displays widespread connectivity with nearly all regions of the brain[49]. As such, it is possible that the heterogeneous encoding of decision variables observed across frontal cortical neurons arises, at least in part, as a result of differences in their downstream connectivity. Given the frontal cortex acts as a key node for the integration and regulation of global brain activity, it seems plausible that the activity of individual frontal cortical neurons encodes information necessary for computations undertaken in the downstream brain regions they project to.

Several studies to date have identified frontal cortical projections that encode specific kinds of behaviorally relevant information[17;105;115]. However, our understanding of how the encoding of specific decision variables maps onto different projection-defined neurons in the frontal cortex - and how this relates to population-level signals - remains limited. Hence, functional analysis of the activity of projection-defined pop-

ulations of frontal neurons during decision-making behaviour could offer important insights into the structure of decision variable representation in the frontal cortex.

Recent work investigating the 'projectome' of frontal cortical neurons in mice has identified the dorsomedial striatum (DMS) and claustrum (CLA) as two major targets of frontal input[48;117]. There is evidence that the DMS contributes directly to decision-making through its role in goal-directed learning and reward-related action selection[119;120;184;185]. While the function of the CLA is less well understood, recent studies have suggested it may contribute to cognitive control, owing to observations of its role in the modulation of frontal cortical activity[135;136] and higher-order cognitive processes such as attention[137;138], engagement[140;141], impulsivity[139], and action control[144–146]. Both of these structures serve as an anatomical 'hub', integrating inputs from many sub-regions of the frontal cortex. The dense inter-connectivity of these two structures with frontal cortical regions, but not with each other, suggests they are likely to be involved in different aspects of frontal cortical function. This makes them good candidates for investigating the relationship between neuronal connectivity and the diverse representations of information in frontal neurons during decision-making behaviour.

This chapter aims to investigate how the representation of decision-relevant information in frontal cortical neurons relates to their downstream projection targets by monitoring the activity of two projection-defined populations of frontal neurons during decision-making behaviour. To this end, we combined large-scale electrophysiological recordings in the frontal cortex of mice performing our visually-guided decision-making task with an in vivo optotagging protocol to examine the neural correlates of decision-making in DMS- and CLA-projecting frontal cortical neurons. This chapter will describe the results of this experiment, outlining how optotagging was implemented to identify DMS- and CLA-projecting neurons and our observations of variable coding in these circuits during decision-making behaviour.

5.2 Results

5.2.1 In vivo optotagging identified populations of DMS and CLA-projecting frontal neurons

An optotagging protocol was implemented to identify frontal neurons projecting to the striatum and/or claustrum (Fig. 5.1A). This method involved delivering a series of targeted laser pulses through optic fibers implanted above the DMS and the CLA to antidromically stimulate the axons of opsin-expressing frontal neurons projecting to these structures (Fig 5.1A). Laser-driven antidromic spikes propagate to the soma, where they can be detected by the Neuropixels probe implanted in the frontal cortex. As a result, this method may be used to separate neurons projecting to the DMS and CLA from the larger dataset of recorded frontal neurons with unknown projection identity. This protocol was carried out immediately after the behavioural task, i.e. when the probe was in the same place but as part of the same electrophysiological recording, in 46 recording sessions across 12 mice. Fig 5.1B shows the histology of one of these mice, confirming that all necessary components (optic fibers, opsin, probe tracts) were correctly targeted.

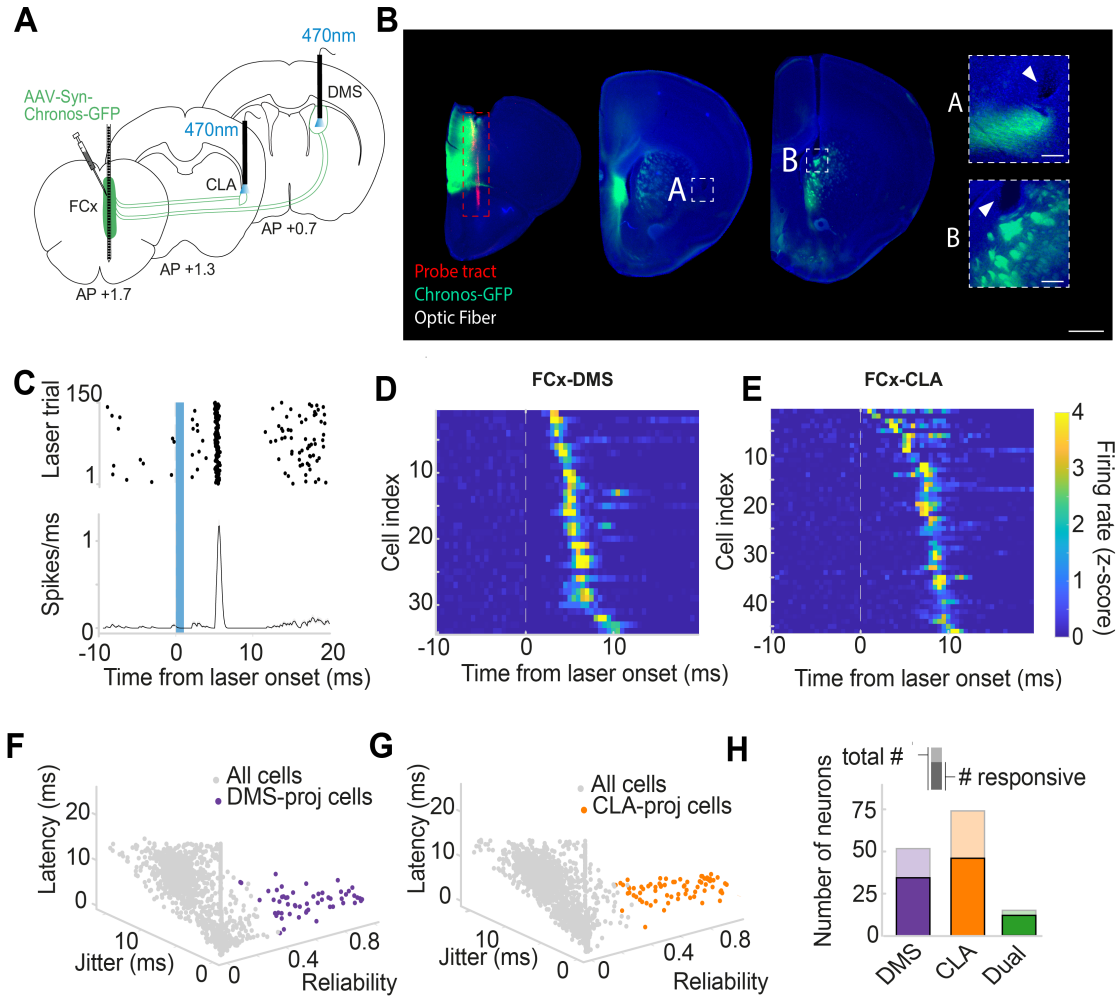


Figure 5.1: In vivo optotagging identified populations of DMS and CLA-projecting neurons in the frontal cortex (FCx) A) Schematic of set up for in vivo optotagging. Chronos-GFP was injected in the frontal cortex and optic fibers implanted above CLA and DMS to enable antidromic axonal stimulation of frontal neurons projecting to these structures via the delivery of targeted laser pules (1ms, 470nm, ˜50mW) B) Histological confirmation that the opsin injection (green), optic fibers (white arrow tips), and Neuropixel probe tract (red) were all correctly targeted, scale bar = 1000µm. A - 10x image of an optic fiber tract above opsin expression in CLA, scale bar = 100µm, B - a 10x image of an optic fiber tract next to opsin expression in DMS, scale bar = 100µm C) Example raster (top) and PSTH (bottom) for one optotagged neuron around laser onset, 1ms laser pulse shown in blue D) Heat-map showing the z-scored firing rate of DMS-projecting frontal neurons after laser onset E) Same for CLA-projecting neurons F) Reliability, latency, and jitter for all recorded frontal neurons, DMS-projecting neurons in purple G) Same for CLA-projecting neurons (orange) H) Total number of optotagged neurons (light colour) and task-responsive optotagged neurons (dark colour) neurons projecting to DMS, CLA, and both (Dual) across all recordings

To identify laser-responsive DMS- and CLA-projecting neurons, all high-quality single units were subjected to rigorous criteria based on existing optotagging literature[17;113–115;186;187]. To be deemed 'optotagged', neurons were required to display a significant change in firing rate relative to pre-stimulus baseline activity within a 10ms window post-laser-onset (stimulus-aligned latency test (SALT), $p < 0.05$; Fig 5.1C), fire a laser-driven spike on $> 40\%$ of laser pulse trials, and display a high correlation ($r > 0.9$) between their spontaneous and laser-evoked waveforms. For more details on the development of this selection criteria, please refer to Appendix B, Section B.6 - Establishing selection criteria to identify laser-responsive neurons.

Implementing these criteria led to the identification of a subset of neurons that displayed robust and short latency responses to laser onset (Fig 5.1D-E), forming separable clusters of laser-responsive neurons relative to the whole recorded population (Fig 5.1F-G). Overall, 66 DMS-projecting neurons and 89 CLA-projecting neurons were identified. Of these, 15 neurons were found to be dual-projecting, meaning their laser-driven responses met the relevant criteria on both DMS and CLA laser stimulation trials. The existence of dual-responsive neurons was notable as no previous study has observed divergent frontal projections to these two structures. An anatomical tracing experiment performed subsequently confirmed the existence of these dual-projecting neurons (see Appendix A.4: Anatomical characterisation of dual-projecting neurons). Dual-projecting neurons were excluded from subsequent functional analyses as the sample size was not sufficient to study them as an independent population of projection-defined frontal cortical neurons.

Consistent with the observations of task-related activity across the frontal cortex, the majority of neurons in both projection-defined populations were found to be task-responsive. 'Task-responsive' was defined using the same metrics as for the whole recorded population across the frontal cortex: displaying a significant difference from baseline firing rate in defined windows around key task events (see Appendix A.3 - Identifying task-responsive neurons). Excluding dual-projecting neurons, 46 CLA-projecting neurons (62.1%) and 33 DMS-projecting neurons (64.7%) were identified as task-responsive (Fig 5.1F).

In summary, an in vivo optotagging protocol was used to identify a subset of projection-defined neurons in the frontal cortex based on their short latency and high-fidelity responses to the delivery of targeted laser pulses. Both CLA- and DMS-projecting neurons were identified, with a small subset of neurons found to project to both structures. The majority of projection-defined neurons were task-responsive. It is these task-responsive frontal neurons - 46 CLA-projecting and 33 DMS-projecting neurons - that will be the focus of analysis in subsequent sections.

5.2.2 Distribution of projection-defined neurons across frontal cortex

Having identified populations of DMS- and CLA-projecting frontal neurons, the first question we sought to address was: how are these projection-defined neurons distributed across the frontal cortex? As in the previous chapter, coating the Neuropixels probes with DiI before recording enabled the location of the probe tracts to be identified during post-hoc histology. Aligning the resulting 3D probe tract reconstructions to the Allen Brain Atlas[152] thus enabled us to map each recorded unit onto its associated brain region.

Region	n DMS-projecting	n total	% projecting
ACA	9	115	7.8
ILA	3	165	1.8
MOs	6	178	3.4
OFC	7	244	2.9
PL	8	291	2.7
Overall	33	998	3.3

Table 5.1: Distribution of DMS-projecting neurons across frontal regions

Overall, 3.4% of task-modulated frontal cortical neurons were identified as DMS-projecting (Table 5.1) and 4.3% as CLA-projecting (Table 5.2) across all sessions in which optotagged neurons were identified. DMS- and CLA-projecting neurons were identified in all frontal regions, with the exception of ILA where no CLA-projecting neurons were identified. The identification of DMS- and CLA-projecting in almost all

regions of the frontal cortex targeted is consistent with recent anatomical work which identified neurons projecting to these structures across all sub-regions and layers of the frontal cortex[48?] . After correcting for the number of neurons recorded per region, a significant over-representation of DMS-projecting neurons was identified in the anterior cingulate cortex (ACA) (Fisher's exact test, $p = 0.003$). Out of 115 task-modulated ACA neurons, 9 (7.8%) were found to be DMS-projecting. This finding is consistent with previous anatomical studies showing a high density of projections from the cingulate cortex to the striatum[188]. No significant over-representation of optotagged neurons was observed in any other frontal region for either DMS- or CLA-projecting populations.

Region	n CLA-projecting	n total	% projecting
ACA	7	131	5.3
ILA	0	181	0.0
MOs	10	141	7.1
OFC	15	257	5.8
PL	14	408	3.4
Overall	46	1118	4.1

Table 5.2: Distribution of CLA-projecting neurons across frontal regions

Thus, examining the distribution of task-modulated CLA- and DMS-projecting neurons identified projection-defined neurons across all regions of frontal cortex, with the significant enrichment of DMS-projecting neurons in ACA.

5.2.3 DMS- and CLA-projecting frontal neurons differentially encode economic variables

The key goal of this chapter is to investigate the encoding of decision variables in projection-defined frontal neurons, and examine how this encoding relates to the patterns of activity observed at a population level across the frontal cortex. Having identified differential patterns of encoding for expected value and risk in the last chapter, we next sought to examine encoding of these variables in neurons projecting specifically to the DMS and CLA.

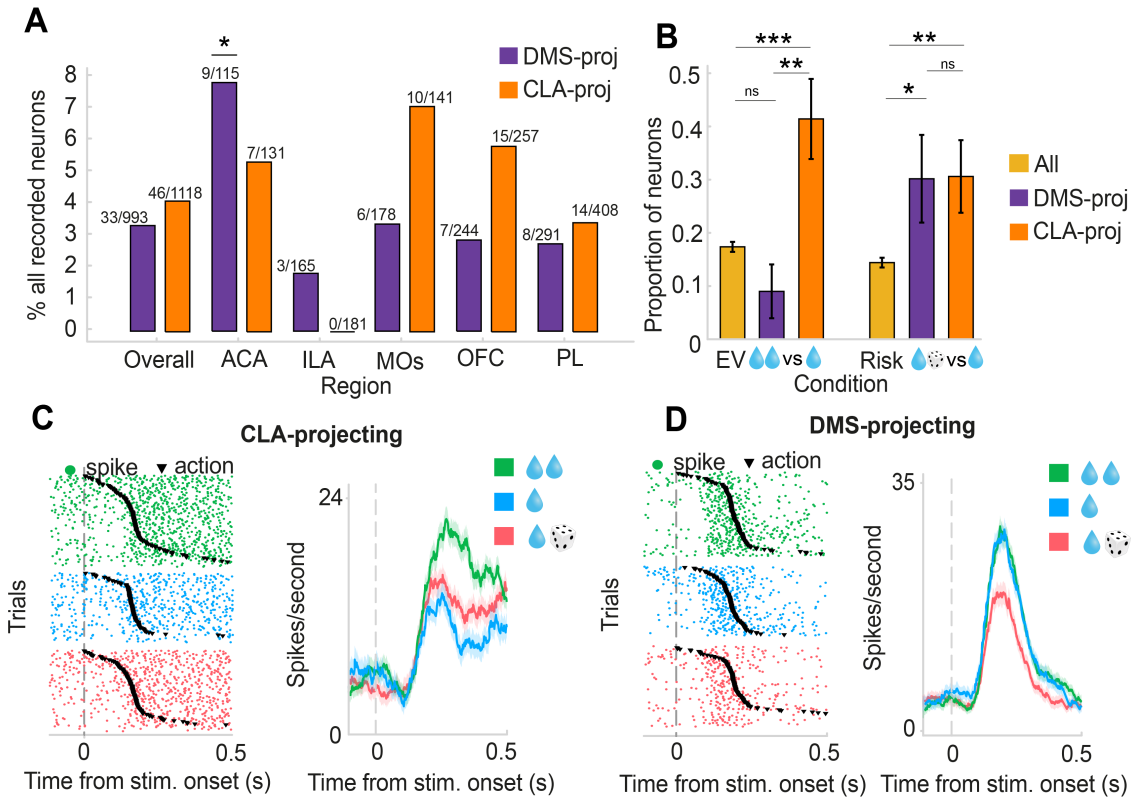


Figure 5.2: Distribution and encoding of DMS- and CLA-projecting frontal cortical neurons A) Proportion of recorded neurons across frontal regions that were found to be DMS-projecting (purple) and CLA-projecting (orange) B) Proportion of neurons encoding EV and risk across all neurons (yellow), DMS-projecting neurons (purple), and CLA-projecting neurons (orange). Errors bars = standard deviation of bootstrapped samples (1000 samples) C) Raster (left) and PSTH (right) for an example CLA-projecting neuron more strongly modulated by the high EV task condition (Large) compared to the low EV conditions (Small and Gamble) D) Example DMS-projecting neurons more strongly modulated by the guaranteed conditions (Large and Small) than the risky condition (Gamble)

The previous chapter employed a linear decoding pipeline to investigate the extent to which expected value and risk could be decoded from the pooled activity of regionally- and spatially-defined populations of neurons across frontal cortex. However, the projection-defined neuronal datasets are considerably smaller. This small sample size makes population decoding methods less reliable and more sensitive to noise, due to the reduced statistical power and increased risk of overfitting. To overcome this difficulty, we implemented a single-cell approach for this dataset; examining differences in trial-averaged post-stimulus firing rates across task conditions per neuron (see Gen-

eral methods and materials, Section 2.2.3- Single cell analysis). This method allowed us to identify neurons with significantly different responses across task conditions in these projection-defined datasets.

Using this approach on the trial-averaged activity of all task-modulated frontal neurons we found that 17.3% (272/1566) of neurons were significantly modulated by EV, while 14.4% (225/1566) of neurons were significantly modulated by risk. This is consistent with the results of the population decoding, which found that population activity more strongly encoded EV than risk, but that both were decodable significantly above chance. Applying this method to the projection-defined data revealed that 41.3% (19/46) of CLA-projecting neurons encoded EV, compared to just 9.1% (3/33) of DMS-projecting neurons. This difference was statistically significant (Z-test, EV-DMS vs EV-CLA, p-value = 0.002). With respect to risk coding, 30.4% (14/46) of CLA-projecting neurons were found to encode risk, along with 30.3% (10/33) of DMS-projecting neurons. This difference was not statistically significant (Z-test, Risk-DMS vs Risk-CLA, p-value = 0.99). Comparing variable coding within projections, a significantly greater proportion of DMS-projecting neurons were risk coding than EV coding (Z-test, EV-DMS vs Risk-DMS, p-value = 0.031), while no significant difference was observed between risk and EV coding for CLA-projecting neurons (Z-test, EV-CLA vs Risk-CLA, p-value = 0.28). EV and risk-coding CLA-projecting neurons appear to be largely not overlapping, with 31.5% (6/19) EV-coding CLA-projecting neurons also coding risk. In aggregate, these results suggest that economic variables are differentially represented in these projections, with a significant proportion of DMS-projecting neurons encoding risk, but not EV, while a significant proportion of CLA-projecting neurons encode both EV and risk.

It is also striking that a larger proportion of projection-defined neurons were found to encode EV and risk compared to the wider population. A significantly higher proportion of both DMS-projecting and CLA-projecting neurons encoded risk relative to all neurons (Z-test, Risk-DMS vs Risk-All, p-value = 0.01; Z-test, Risk-CLA v Risk-ALL, p-value = 0.003), while a significantly higher proportion of CLA-projecting

neurons encoded EV compared to all neurons (Z-test, EV-CLA vs EV-All, p-value $<$ 0.00001).

In summary, a single-neuron approach examining differences in trial-averaged firing rate across different trial conditions revealed that a higher proportion of projection-defined neurons encoded EV and risk compared to the overall population. These results also point towards a difference in EV and risk coding between CLA-projecting and DMS-projecting frontal cortical neurons, with DMS-projecting neurons more strongly encoding risk, while a largely non-overlapping subset of CLA-projecting neurons encoded both variables.

5.2.4 Differential encoding of EV and Risk in DMS- and CLA-projecting neurons in dorsomedial frontal cortex

In the previous chapter, we identified distinct spatial gradients for the encoding of risk and EV across the frontal cortex, with stronger coding of risk observed in medial but not lateral frontal areas and stronger coding of EV in dorsal but not ventral areas. We were therefore curious to test whether the differential coding observed in DMS and CLA projections also reflected these spatial gradients, or whether it was independent of the frontal region in which these neurons originate.

To address this question, we compared the coding of EV and risk in projection-defined neurons originating specifically in dorsomedial regions of frontal cortex (ACA and PL). In our population decoding analysis, these areas displayed strong coding of both EV and risk. We therefore sought to test if the coding differences between DMS- and CLA-projecting neurons persisted when focusing on neurons from these dorsomedial regions. If these differences arose from a sampling bias in specific locations along the EV and risk coding gradients, we would expect less of a difference between projections originating in dorsomedial frontal cortex (dmFCx), where strong coding of both variables was observed. Neurons from dmFCx represented approximately half of the projection-defined dataset: 51.5% (17/33) for DMS-projections and 45.6% (21/46) for CLA-projections.

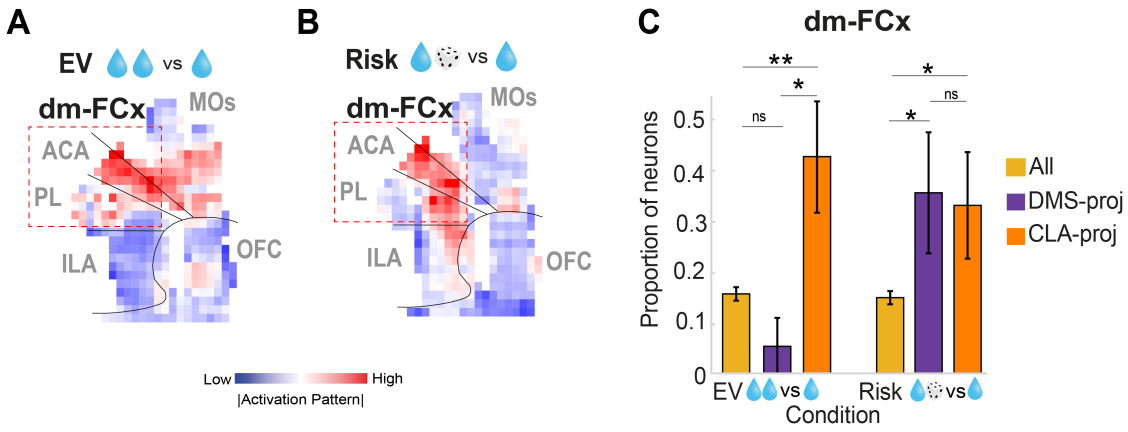


Figure 5.3: Encoding in dorsomedial frontal cortex neurons projecting to DMS and CLA A) The dorsal-to-ventral gradient observed for EV at a population-level, red box outlines dorsomedial regions (ACA and PL) B) The medial-to-lateral gradient observed for risk at a population-level, red box outlines dorsomedial regions (ACA and PL) C) Proportion of dorsomedial frontal neurons encoding EV and risk across all neurons (yellow), DMS-projecting neurons (purple), and CLA-projecting neurons (orange). Errors bars = standard deviation of bootstrapped samples (1000 samples)

Figure 5.3C shows the proportions of dmFCx neurons encoding EV and risk. Remarkably, the same coding differences were observed in this spatially restricted set of neurons compared to the entire dataset. A higher proportion of dmFCx neurons projecting to DMS encoded risk (35.2%, 6/17) compared to EV (5.9%, 1/17), while dmFCx neurons projecting to CLA encoded both EV (42.9%, 9/21) and risk (33.3% 7/21). These proportions were significantly higher than the proportion of EV and risk coding neurons identified in all recorded dmFCx neurons with unknown projection identity (EV: 16.8%, 123/734; risk: 15.9%, 117/734; Z-tests: EV-CLA-dmFCx v EV-all-dmFCx, $p = 0.002$; Risk-CLA-dmFCx v Risk-all-dmFCx, $p = 0.034$; Risk-DMS-dmFCx v Risk-all-dmFCx, $p = 0.033$), with the exception of EV coding in DMS-projecting neurons which was not significantly different (Z-test, EV-DMS-dmFCx v EV-all-dmFCx, $p = 0.258$). Thus, projection-defined neurons originating in an overlapping area of strong EV and risk coding - dorsomedial regions of frontal cortex - still displayed significant differences in their encoding of these variables, both relative to the wider population of dmFCx neurons and to each other.

5.3 Discussion

This experiment sought to develop our understanding of the neural correlates of decision-making across the regions and projections of the frontal cortex by integrating large-scale electrophysiological recordings with an in vivo optotagging protocol to monitor the activity of projection-defined frontal neurons in mice performing a decision-making task. Specifically, this experiment focused on the encoding of decision variables in frontal neurons projecting to the dorsomedial striatum and the claustrum.

In vivo optotagging successfully identified a subset of DMS- and CLA-projecting frontal neurons that displayed short latency and high-reliability responses to laser-onset. A subset of neurons were responsive to both DMS and CLA stimulation, indicating these neurons send divergent axon collaterals to both structures. As divergent frontal projections to DMS and CLA had not previously been reported, an anatomical tracing experiment was performed to confirm the existence of dual-projecting neurons. The majority of neurons in all three groups (DMS-, CLA-, and dual-projecting) were found to be task-responsive.

Examining the distribution of DMS- and CLA-projecting neurons across frontal regions revealed an over-representation of DMS-projecting neurons in anterior cingulate cortex. There was not a significant over-representation of optotagged neurons in any other region of frontal cortex, although both DMS- and CLA-projecting neurons were identified in all regions targeted. Previous anatomical characterisation of corticostriatal projections suggests that one would expect to observe striatal-projecting neurons across frontal cortical regions, but particularly in anterior cingulate cortex, secondary motor cortex, and orbitofrontal cortex[188]. While the enrichment of ACC-DMS projections is most evident in these data, DMS-projections were indeed observed across all these expected regions.

Several anatomical tracing studies have observed that the CLA also receives dense input from anterior cingulate cortex[48;117;144]. Indeed, the ACA-CLA pathway has

previously been implicated in top-down control of goal-directed action as disrupting its activity has been found to inhibit performance in a 5-choice serial reaction time task[144]. CLA-projecting ACA neurons were identified in this dataset (7/46), but no more so than in any other frontal region. However, it is worth noting that anterior cingulate input to claustrum is predominantly contralateral - while this optotagging set-up solely targeted ipsilateral projections. The rationale for this ipsilateral focus was that previous tracing of frontal cortical inputs to claustrum, aside from those originating in ACA, were found to be largely ipsilatera;[117]. Thus, to specifically examine the bulk of anterior cingulate input to claustrum in the context of decision-making behaviour, this experiment would likely need to be repeated with opsin expression and in vivo recordings in the contralateral hemisphere to CLA stimulation. This was not feasible in the time frame permitted for this set of experiments but represents a useful goal for future work.

The primary goal of this chapter was to examine the encoding of key decision variables — namely expected value and risk — in projection-defined frontal neurons and to compare these findings with the patterns of encoding observed across the frontal cortex in the previous chapter. We observed that a higher proportion of CLA- and DMS-projecting neurons encoded EV and risk, respectively, relative to all recorded neurons. This analysis also revealed a difference in encoding between DMS- and CLA-projecting neurons, as DMS-projecting neurons more strongly encoded risk, while CLA-projecting neurons were found to encode both variables. The representation of both EV and risk, including a particularly strong coding of EV in CLA-projecting neurons was striking as this frontal projection has not previously been studied in the context of value-based decision-making. The strong representation of economic variables observed in this study indicates that the CLA may play a previously overlooked role in the neural computations that underpin decision-making, although follow-up work will be required to examine any causal effects of these representations on behaviour.

Notably, we did not observe the same over-representation of expected value coding in striatum-projecting OFC neurons reported by Hirokawa et al (2019)[115]. However,

one potential explanation for these functional differences could be the specific neuronal populations targeted by our respective methods. In their study, Hirokawa et al targeted lateral OFC neuron, while this work largely targeted medial OFC neurons. A recent study[48] examining the projection targets of PFC neurons demonstrated that specific sub-types of PFC neurons project to distinct areas of the striatum, with neurons originating in ventral OFC and medial OFC projecting largely to ventral and medial striatum, respectively. Thus, the functional difference observed between striatum-projecting OFC neurons in this study compared to Hirokawa et al may be attributed to targeting different neuronal populations with distinct striatal projection patterns.

Finally, a control analysis was conducted to determine whether the observed differences in encoding between DMS- and CLA-projecting neurons could be attributed to the same spatial gradients identified across frontal sub-regions in Chapter 4. Specifically, we examined the coding of expected value and risk in neurons projecting from the dorsomedial regions of the frontal cortex, namely the anterior cingulate and prelimbic cortex. Our population decoding analysis showed strong coding of both EV and risk in these areas. We therefore hypothesised that if the spatial gradients across sub-regions were driving the observed differences between projections, then focusing on neurons from these dorsomedial regions would eliminate or reduce the coding differences between DMS- and CLA-projecting neurons. However, our analysis revealed that the coding differences persisted even when considering only dorsomedial projection-defined neurons. This indicates that the differences in EV and risk coding between DMS- and CLA-projecting neurons are not simply reflective of the spatial gradients observed at a population-level within the frontal cortex.

Taken together, the results of this experiment suggest that different populations of projection-defined neurons encode economic variables beyond the spatial origins of those projections. Although this may seem counter-intuitive, it points to the existence of decision variable representation within the frontal cortex operating on two distinct scales. In other words, while the heterogeneous variable encoding observed across frontal neurons may be partly explained by focusing on specific projection pathways,

projection-specific coding may still differ from that observed across the population as a whole. These observations underscore the necessity for further research into how different populations of projection-defined neurons integrate to form the structured representations observed at the population level. Furthermore, this finding reinforces the importance of taking a large-scale and dynamic view of brain function. Isolated analyses of specific regions or projections will not suffice to fully understand the collaborative dynamics of neural circuits underlying complex behaviours. Rather, a comprehensive, integrative approach that examines variable representation across spatial scales is essential for advancing our understanding of the neural mechanisms that underpin decision-making.

5.4 Limitations and future directions

Perhaps the most significant limitation of implementing in vivo optotagging in this experiment is that it is inherently low throughput. The frontal cortex sends input to many downstream structures meaning the probability of any given recorded neuron projecting to DMS and/or CLA is low. Although these structures were chosen in part for their dense frontal cortical input, resulting in identified optotagged cells in most recordings, only a small number of laser-responsive neurons were identified per mouse. Combined with the time taken to train animals in this task, and the fact that one cannot verify all necessary components (opsin injection, fiber implantations, and Neuropixels probe) were appropriately targeted until post-hoc histology at the end of recording, generating this kind of projection-defined neuronal dataset is both low throughput and time-intensive.

In addition, optotagging sessions were subject to the same behavioural and electrophysiological inclusion criteria applied to all sessions, such that a valid comparison could be made between the projection- and population-level datasets. This led to the exclusion of several recording sessions containing optotagged neurons, primarily because they did not meet the minimum number of trials per stimulus condition required for population decoding. Given the relative size of these two datasets, these

strict criteria will have resulted in a proportionally greater loss of data from the optotagging dataset. Thus, in addition to the inherently low throughput nature of optotagging as a technique, throughput was further reduced by steps taken to ensure compatibility between the two datasets.

The key consequence of this low throughput is that it makes it challenging to compare encoding in projection-defined neurons originating in different frontal sub-regions, or to create the same kind of coding map across the frontal cortex as was possible for the population data. It is therefore possible that more extensive optotagging data may reveal different relationships between coding in these projection-defined neuronal populations, the frontal regions in which they originate, and the coding gradients observed across the frontal cortex as a whole. Nonetheless, acknowledging that this dataset is comparable in size to other published work using in vivo optotagging[17;115], the results of this experiment point towards differential representation of economic variables in DMS and CLA-projecting frontal neurons.

Future work could further examine the contribution of these projection-specific representations to decision-making behaviour through causal manipulations, or through the adaptation of the behavioural task to examine how a wider variety of stimuli corresponding to different levels of expected value and risk affects coding in these neuronal populations. Similarly to the population data, the current analysis outlined above focuses specifically on post-stimlus, pre-reward responses on correct trials. Additional analyses could therefore also examine projection-specific activity in different epochs of the task - e.g outcome evaluation, error trials, engaged and disengaged states - to examine variable representation in these projection pathways during distinct features of decision-making behaviour.

5.5 Summary

The goal of this experiment was to examine decision variable encoding in projection-defined frontal neurons projecting to the dorsomedial striatum and the claustrum.

Combining in vivo optotagging with large-scale electrophysiological recordings in mice performing a decision-making task enabled us to identify a subset of task-modulated frontal neurons projecting to these downstream structures. Examining EV and risk coding across projection-defined neurons revealed these cells are differentially recruited during decision-making behaviour: DMS-projecting neurons display stronger risk coding, while CLA-projecting neurons were found to encode both EV and risk. Notably, a higher proportion of these neurons were found to encode these decision variables compared to the wider population. These results indicate that the response properties of frontal cortical neurons may be better resolved by studying specific downstream projections, and allude to different roles for corticostriatal and corticoclaustral projections in decision-making behaviour.

Chapter 6

Discussion

Economic decision-making involves evaluating different options and their associated outcomes to select actions that align with our goals. Extensive research has indicated that the frontal cortex plays a crucial role in the neural computations underlying economic decision-making. Lesion studies[5–11] and neuroimaging work[53;54;61;174;175] have provided evidence that specific structures in the frontal cortex may support distinct features of economic decision-making, while electrophysiological approaches have identified populations of neurons within these regions that encode key economic variables such as expected value, expected outcome, reward probability, and choice[12–17].

However, the development of novel tools for monitoring neuronal dynamics with better temporal and spatial resolution across larger scales has uncovered new layers of complexity in the neural basis of economic decision-making. In particular, neurons across different frontal regions have been found to heterogeneously encode an overlapping array of variables, indicating that the encoding of decision-relevant information may be more widely distributed in frontal cortical circuits than previously assumed[85;104]. Furthermore, a growing body of work has provided evidence that different neuronal cell types—which can be variously defined by spatial location, genetic identity, biophysical properties, and connectivity—may encode specific kinds of information[105;112;115;189] and shape the flow of this information through the brain in different behavioural contexts[113]. As such, it remains unclear how economic variables are represented across the frontal cortex and how these representations are integrated

to support decision-making and goal-directed behaviour.

The key question this thesis sought to address is: how are the neural correlates of economic decision-making structured in the regions and projections of the frontal cortex? To achieve this, we developed a preparation that allowed us to investigate the encoding of specific economic variables across region- and projection-defined frontal neurons during decision-making behaviour in mice.

6.1 Economic decision-making in mice

We established a novel visually-guided economic decision-making task to study decision-making behaviour in mice. In this task, head-fixed mice were required to make choices in response to visual stimuli predictive of different outcomes, with reward magnitude and probability systematically manipulated between trials. While several value-based decision-making paradigms have been reported in previous rodent studies[163;164], our task was designed to be more akin to those used in primate literature, requiring mice to make choices in response to abstract visual stimuli whose sensory features do not correlate with the rewards they represent. This simplifies the interpretation of the resulting neural activity, as the neural encoding of economic features of the presented options will be less confounded by highly correlated sensory signals. We demonstrated that mice were able to make choices in response to these abstract visual stimuli with a high degree of accuracy, and observed that their accuracy scaled with the expected value of the presented options. These results suggest that mice, like humans and non-human primates, can integrate visual information and past experiences to guide choices in a manner consistent with existing economic decision-making theories.

6.2 Spatial gradients in neural encoding of economic variables

Large-scale electrophysiological recordings were then used to monitor the activity of almost 7,000 frontal neurons in mice performing this task. The goal of this experiment was to examine how the fundamental economic variables manipulated in the task — expected value (EV) and risk — were encoded across the various sub-regions of the frontal cortex. Using a linear decoding approach, we observed that the pooled activity of all recorded frontal neurons reflected differences in expected value, with significant encoding of both EV and risk observed across all frontal regions studied. This suggests that frontal regions do not display clear qualitative differences in the encoding of certain economic variables, but that quantitative differences may still be identified by comparing the extent of encoding across regions. This echoes the findings of other studies comparing the coding of economic variables in different regions of the frontal cortex[85;104], which have observed widespread coding of multiple decision variables, but with significant differences in the nature of this coding between regions. Notably, our results did not highlight the same dominance of the orbitofrontal cortex (OFC) in EV coding that has been widely reported in some of the most seminal papers in neuroeconomics[12;13;89]. Rather, dorsal regions such as the anterior cingulate cortex (ACC) and secondary motor cortex (MOs) were identified as displaying particularly strong coding of EV. While the structure of this task and the use of over-trained mice may offer an alternative explanation for the lack of OFC engagement in this context[179], these results nonetheless suggest that other structures in the frontal cortex may play an important and somewhat overlooked role in the encoding of EV in the context of economic decision-making.

One of the most notable results from comparing the encoding of EV and risk across regions is the strong coding of both variables observed in the anterior cingulate cortex. Value and risk signals have frequently been identified in the ACC, along with a wide variety of other decision variables[85;190;191]. Heilbronner and Hayden (2016)[182] proposed that the ACC may contribute to decision-making by "linking contexts with

strategies [and] integrating diverse task-relevant information to create a rich representation of task space." Related to this, it has also been proposed that the ACC represents a key node for the allocation of control in decision-making[192], particularly with respect to the "expected value of control." This theory offers a way of reconciling the ACC's observed role in representing value, risk, effort, conflict, and outcome by suggesting that the function of the ACC is to generate a comprehensive representation of task space and integrate that information to allocate cognitive resources and drive goal-directed action. Our results further support the idea that the ACC is a key nexus for encoding decision-relevant information. However, additional research is necessary to examine the behavioural consequences of disrupting these representations to clarify their specific contribution to economic decision-making behaviour.

Examining the spatial structure of EV and risk coding across the frontal cortex further revealed distinct spatial gradients underlying the representation of these variables, with EV displaying a dorsal-to-ventral gradient and risk a medial-to-lateral gradient. These coding gradients were largely driven by trials in which stimuli were presented contralaterally to the recorded hemisphere. It is worth noting that this is not the first study to report the existence of coding gradients in the context of decision-making. For example, Maisson et al. (2021)[104] trained macaques in a risky decision-making task and observed a functional coding gradient of choice information along the dorsoventral axis of the medial prefrontal cortex. In particular, they observed that decodability of two economic variables — expected value and chosen offer — increased in the ventral-to-dorsal direction, with the highest decodability observed in the dorsal anterior cingulate cortex. The authors suggest that this increased decodability may represent the progressive transformation of information about options into choices, with ventral regions serving as the entry point for this information and dorsal regions the top of the hierarchy, responsible for relaying key information to output structures. The increased decodability in dorsal compared to ventral regions reported in this paper aligns with the dorsal-to-ventral EV gradient observed in our study. Importantly, we additionally show that a distinct medial-to-lateral gradient exists for risk coding. These observations support our findings of spatial gradients

underlying the coding of economic variables in the frontal cortex. Spatial gradients have also been reported in other brain regions in different behavioural contexts[193;194], suggesting that this form of spatially structured representation is not specific to economic decision-making. Rather, such gradients may represent a common framework for the neural representation of task-relevant information, although further research is needed to confirm this.

6.3 Projection-specific coding

Finally, we used in vivo optotagging to monitor the activity of frontal neurons projecting specifically to the dorsomedial striatum (DMS) and the claustrum (CLA). These structures receive dense, integrative input from multiple frontal cortical regions and have been implicated in goal-directed action selection and cognitive control, respectively. We observed that a higher proportion of these projection-defined neurons encoded expected value (EV) and risk compared to the general population. Notably, DMS-projecting neurons primarily encoded risk, while CLA-projecting neurons encoded both EV and risk. Specific encoding of risk in frontal inputs to the DMS had not been previously reported, although there is evidence that the DMS may facilitate flexible action selection in the context of risky decision-making. For instance, Schumacher et al. (2021)[122] demonstrated that inactivating the DMS impaired rats' ability to adjust choice biases following changes in risk in a probabilistic discounting task. Additionally, D2/3 receptor binding within the dorsal striatum has been found to correlate with increased risk-taking in risky decision-making tasks[195]. It is possible that frontal inputs to the DMS may support dynamic responses to changing reward contingencies by signalling the probability of rewarded outcomes, although this remains speculative as our data cannot provide insight into whether the observed signals play a correlative or causal role in choice behaviour under uncertainty. An adaptation of the behavioural task to focus on risky decision-making—such as the inclusion of stimuli corresponding to a range of reward probabilities or choices between guaranteed and risky rewards—is likely required to more comprehensively investigate

the encoding of risk in this neural circuit.

The observation that CLA-projecting neurons encode both risk and EV was also striking, particularly as frontal inputs to the CLA have not previously been studied in the context of economic decision-making. Previous work has suggested that the CLA may play a role in cognitive and attentional control, likely through its modulatory effect on frontal cortical circuits[135]. For example, Atlan et al. (2018)[140] found that inhibiting CLA neurons impacted the performance of mice on an automated two-alternative forced-choice (2AFC) behavioural task upon the introduction of a distractor, suggesting a specific role for the claustrum in supporting the maintenance of goal-directed attention. It therefore seems plausible that frontal inputs to the claustrum may provide a comprehensive overview of task variables to support CLA-mediated allocation of attentional resources. However, further work examining the behavioural effects of disrupting these pathways will be required to test hypotheses about their respective roles in decision-making behaviour.

Overall, our results suggest potentially distinct roles for corticostriatal and corticoclaustral projections in economic decision-making. Furthermore, this provides evidence that the encoding of decision variables in the frontal cortex not only follows a spatial organisation, but is also organised according to projections to downstream brain regions.

6.4 Future directions

The results of this study open the door to several streams of future research. For instance, further work involving experimental manipulations is required to determine the extent to which the representations uncovered in this study play a causal role in economic decision-making behaviour. Based on the spatial coding gradients we observed at a population level, we hypothesise that disrupting dorsal frontal regions — perhaps by using optogenetic or chemogenetic approaches — would likely affect behaviours reliant on the computation of value, while disrupting medial regions would

exert more of an effect on risk-based decision-making. This disruption may be specific to information presented contralateral to the manipulated neural circuits, given the differential activation patterns we observed for ipsilateral and contralateral trials in this study.

The behavioural task could also be adapted to explore other aspects of economic decision-making not captured in this study. For example, mice could be trained to respond to auditory or olfactory stimuli to determine whether the spatial gradients of EV and risk coding observed in Chapter 4 generalise across sensory modalities, or if they are specific to visually-guided economic choices. Other adaptations could involve the inclusion of more stimuli pertaining to a wider range of potential reward magnitudes and probabilities, and choices between stimuli, to more comprehensively probe subjective preference and risk attitudes, and the neural encoding thereof.

Finally, it is valuable to consider the potential utility of these findings in the context of human studies and eventual clinical applications. Biased or disordered decision-making is a feature of several major psychiatric disorders, including addiction[32;33], obsessive-compulsive disorder (OCD)[34], schizophrenia[35], and depression[36]. The observed link between certain pathologies and abnormal decision-making processes suggests that gaining a deeper understanding of how decision variables, such as expected reward (e.g., the experience of drug-taking) and risk (e.g., relapse, health implications, socioeconomic consequences), are represented and altered in psychiatric disorders could help us better understand and support individuals managing these conditions, and perhaps identify novel targets for treatment further down the line. Future research could therefore seek to develop human economic decision-making paradigms similar to the one used in this study to test if the representational gradients observed here are also a feature of frontal cortical function in humans and examine how these representations are altered in patients with disorders linked to impaired decision-making.

6.5 Conclusion

In summary, this thesis used a combination of behavioural, electrophysiological, and optogenetic approaches to study the structure of economic variable representation in the regions and projections of the frontal cortex. We revealed a novel spatial structure underlying the encoding of expected value and risk across the frontal cortex, and provided evidence that economic variables are differentially represented in corticostriatal and corticoclaustral circuits. We conclude that the encoding of economic variables spans multiple scales of neural representation in the frontal cortex, and suggest that future experiments seeking to uncover the neural basis of decision-making should endeavour to operate across these scales.

Appendix A

Appendix A

A.1 Two-stimulus trials

To provide further evidence that mice are sensitive to differences in the expected value of the stimuli presented in this task (see Chapter 4 for details), a subset of animals (n = 13) were trained in a separate study on 'two-stimulus trials'. In this experiment, mice were presented with two visual stimuli simultaneously and were required to choose between them. This task did not use a block structure: each stimulus could appear on either the right or left-hand side of the screen in every trial, meaning mice were required to make a trial-by-trial assessment of the location of their preferred stimulus.

There were two conditions in which mice were presented with two stimuli that predicted outcomes with different expected values: the Large versus Small condition and the Large versus Gamble condition. Consistent with their performance in single stimulus trials, expert animals presented with a choice between visual stimuli corresponding to different expected values consistently selected the higher value (Large) stimulus (see Fig A.1A and Table A.1). No significant difference in performance was observed in Large versus Gamble trials compared to Large versus Small trials. Performance was notably lower overall on the two stimulus task compared to the single stimulus task, likely due to the additional attentional and cognitive demands of

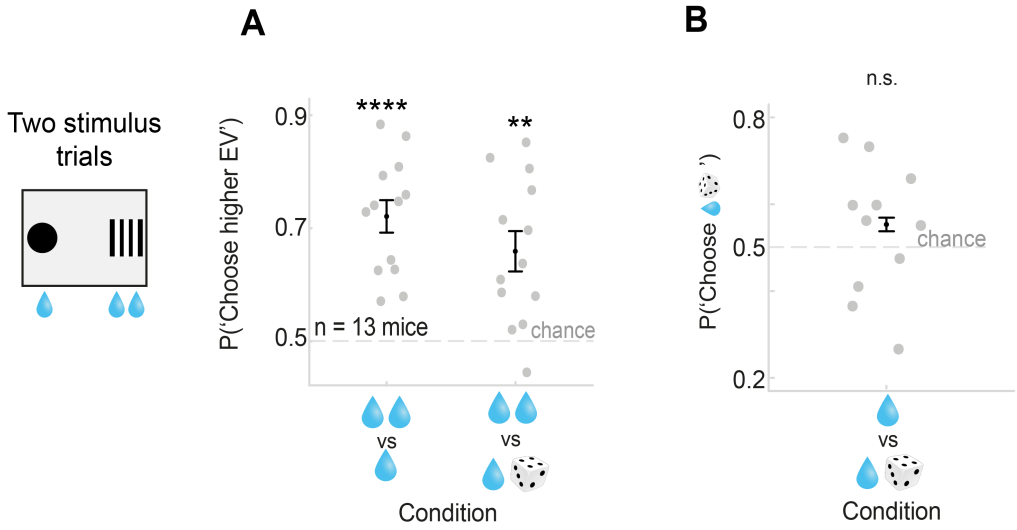


Figure A.1: Performance on two stimulus trials A) Accuracy on two stimulus trials (n = 13 mice). Mice trained on two stimulus trials consistently selected the higher value stimulus when presented with options that differ in expected value (i.e Large vs Small and Large vs Gamble) B) There was no significant difference in preference for stimuli with the same EV (Small vs Gamble)

Condition	p('Chooses higher EV')	s.e.m	p-value (against chance)
Large v Small	0.721	0.0287	$5.49e^{-}6$, ****
Large v Gamble	0.866	0.0357	$7.64e^{-}5$, ****

Table A.1: Mean accuracy across two stimuli trial conditions (different EV)

having to process and compare two reward-predictive stimuli (compared to a choice between a reward-predictive stimulus and nothing). Nonetheless, the significantly greater preference for higher-value stimuli observed across mice when offered a choice between stimuli predicting a high- or low-value outcome provides additional evidence that mice are sensitive to value differences and can make value-based decisions to maximise reward.

There was also a two-stimulus trial condition in which the two stimuli presented had identical expected values but differed in their associated risk (Small v Gamble). In this case, mice did not display a significant preference for either the Small or Gamble stimulus. Notably, mice varied in which of these two stimuli they preferred: some

Condition	p('Chooses Gamble')	s.e.m	p-value (against chance)
Small v Gamble	0.553	0.015	0.552, ns

Table A.2: Mean accuracy across two stimuli trial conditions (samet EV, different risk)

exhibited a pronounced preference for the Gamble stimulus, some favoured the Small stimulus, and a third group showed no marked preference for either. While this may allude to differences in risk preference across mice, it is important to note that a thorough exploration of risk preference would necessitate the use of a broader array of stimuli varying in reward magnitude and probability. This approach would allow for the construction of a subjective utility curve, which could quantitatively represent the extent to which individual animals are willing to engage in risky behaviours in exchange for potential rewards. Risky behaviour cannot be comprehensively probed from the current experimental setup, however, as only one uncertain option was presented in this task. Nonetheless, this could represent a potential adaptation of the behavioural paradigm for future experiments.

A.2 Identifying task-responsive neurons

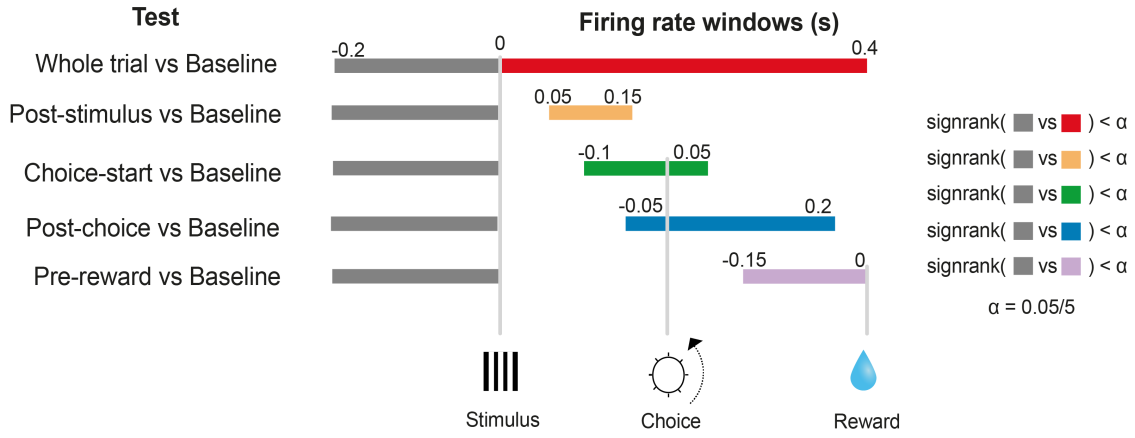


Figure A.2: Identifying task-driven responses across task epochs. The trial-averaged firing rate was compared to pre-stimulus baseline firing rate in four task windows: whole trial, post-stimulus, choice-start, post-choice, and pre-reward. Reward-related activity was excluded from this analysis. Signed rank tests corrected for multiple comparisons (alpha = 0.05/5) were used to test the significance of these firing rate differences

To identify 'task-responsive' neurons, trial-averaged firing rate was compared to pre-stimulus baseline firing rate in several time windows around key task events for every neuron. Neurons were deemed task-responsive if they displayed a significant deviation from baseline firing rate in one or more of these task windows (Wilcoxon sign rank test with Bonferroni correction, $p < 0.01$).

A.3 Correlating variable encoding

One of the key results detailed in Chapter 4 is that EV and risk coding exhibit distinct spatial gradients across the frontal cortex, largely driven by responses elicited by contralateral stimulus presentation. An outstanding question regarding these representations is: are these variables encoded by overlapping or independent populations of frontal neurons? The theory of mixed selectivity[196] suggests that neurons, particularly neurons in the frontal cortex, can display selectivity for multiple variables at once. This suggests that we may expect to find neurons representing multiple decision variables simultaneously. However, the observation that EV and risk display distinct spatial gradients indicates that a significant portion of neurons do not encode both variables, as one would expect an entirely overlapping population of EV and risk-coding neurons to generate the same spatial pattern of activation. Stimulus location, on the other hand, appears to display a similar pattern of activation across the frontal cortex as EV, although the 2D activation maps provide no insight into whether this pattern of activation arises from the same or different groups of neurons in the regions. Thus, it remains unclear the extent to which neurons encoding EV, risk, and stimulus location overlap.

To address this question, we examined the correlation between the absolute activation weights for each pair of variables (EV, risk, and stimulus location) for all neurons (Fig A.3, top row). As a reminder, the magnitude of the activation weight indicates the extent to which neuronal responses can distinguish between two different stimulus conditions (high EV vs low EV, ipsi vs contra etc). Interestingly, the correlations between stimulus location activation weights and both EV and risk activation weights

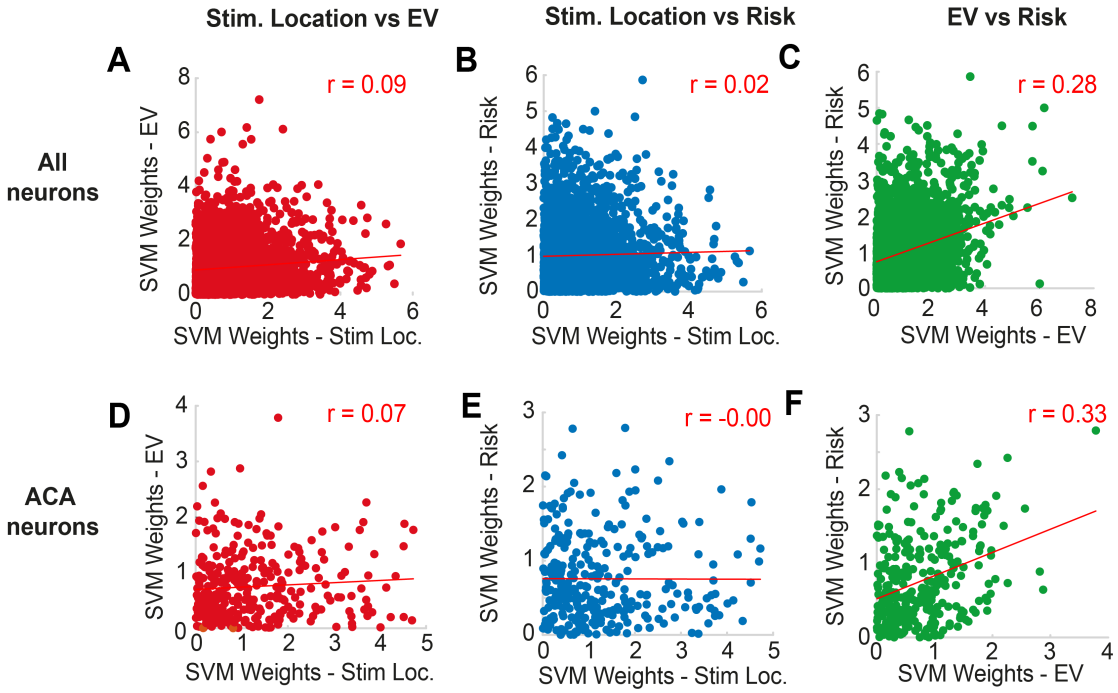


Figure A.3: Stimulus location, EV, and risk coding correlation. Absolute SVM weight correlation between stimulus location and EV (A + D), stimulus location and risk (B + E), and EV and risk (C + F) for all neurons (top row) and ACA neurons (bottom row)

were very weak (stimulus location vs EV, r = 0.09, stimulus location vs risk, r = 0.02). In contrast, a moderate correlation was observed between EV and risk (r = 0.28). This suggests that, across all recorded neurons, there was no clear relationship between the coding of stimulus location and the outcomes those stimuli represented. Neurons were more likely to encode both EV and risk than either of these variables and stimulus location.

Given this initial correlation was performed on all recorded neurons, it remained a possibility that neurons in specific regions of the frontal cortex could be more highly correlated than the global correlations suggested. This question is of particular interest given the strongly overlapping areas of encoding observed for stimulus location and EV, both of which display the greatest encoding in dorsal regions of the frontal cortex. This suggests that we might expect to observe a higher correlation between stimulus location and EV in dorsal areas, such as the anterior cingulate cortex (ACA).

However, examining the correlations between EV, risk, and stimulus location specifically in ACA neurons did not reveal a significantly greater degree of overlap for any of these variables (Fig A.3, bottom row). Consistent with the correlations across the whole population, the correlation between stimulus location and the other two variables was weak (r = 0.07 and r = -0.00), while a moderate correlation was observed between EV and risk (r = 0.33).

Thus, despite the overlapping regions of activation observed in dorsal regions for the encoding of EV and stimulus location: this coding appears to arise from largely independent populations of neurons. There does not appear to be a pronounced overlap in the encoding of stimulus location and expected value or risk, both across the whole population or in the ACA. Similarly, the moderate correlation observed between EV and risk weights in ACA neurons was not notably different to the whole population. As such, while the weak-to-moderate positive correlations identified in this analysis indicate a degree of mixed selectivity for the encoding of these variables in the frontal cortex, it nonetheless suggests that the encoding of stimulus location, EV, and risk are largely categorical.

A.4 Anatomical characterisation of dual-projecting neurons

The identification of dual laser-responsive neurons described in Chapter 5 alludes to the existence of neurons that send divergent axon collaterals to both DMS and CLA. Previous tracing work has identified dense projections from frontal cortical regions to the DMS and CLA[48;117], however, no studies have tested whether a subset of frontal cortical neurons sends input to both. Seeking anatomical confirmation of these electrophysiologically identified dual-projecting cells, we conducted a tracing experiment to examine if frontal neurons projecting to DMS also project to CLA. Briefly, retro-Cre was injected into the DMS and a Cre-dependent fluorescent marker (Syn-FLEX-ChrimsonR-tdTomato) in the frontal cortex - using the same injection coordinates as for the broader optotagging experiment (see General methods and

materials, Section 2.2.3 - Intracranial injections and optic fiber implantations). This conditional expression system should only result in visible expression in CLA if there are DMS-projecting frontal cortical neurons that also send axon collaterals to CLA.

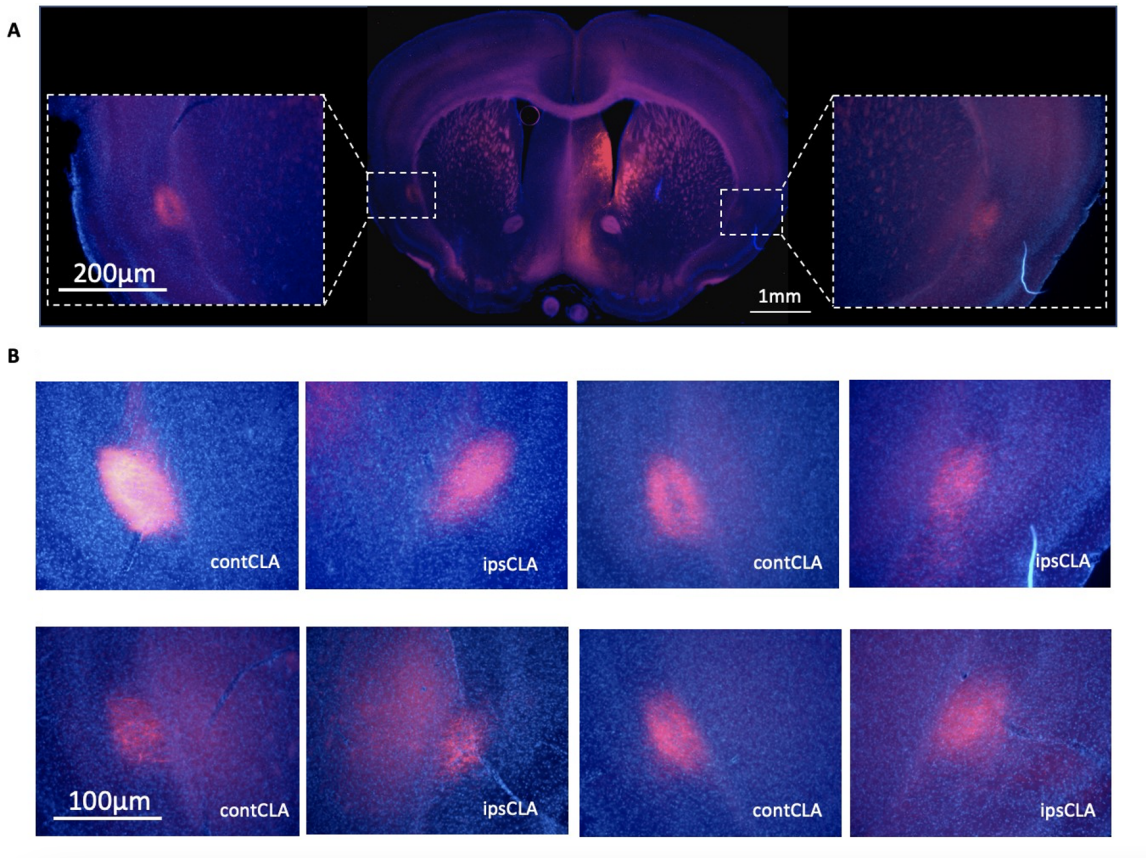


Figure A.4: Anatomical tracing confirmed the existence of divergent frontal projections to DMS and CLA A) Histology from one mouse in which TdTomato expression was observed in contralateral and ipsilateral CLA following Cre-dependent expression of the opsin in DMS-projecting frontal neurons B) 10x images of opsin expression in contralateral and ipsilateral CLA for all animals used in this tracing experiment.

Following these tracing surgeries, once sufficient time had elapsed for visible viral expression to develop, the animals were sacrificed and brain slices prepared for imaging. An ImageJ workflow was developed to compare tdTomato expression in the CLA to fluorescence at a) the injection site (DMS) and b) background fluorescence. The purpose of these comparisons was to determine whether tdTomato expression observed in CLA was significantly greater than background fluorescence. Normalising measures of fluorescence to expression observed at the injection site accounted for differences in

viral expression across mice. Across all injected mice (n = 4), a significant difference was observed between background fluorescence and fluorescence of the contralateral CLA and ipsilateral CLA (paired t-tests, $p < 0.008$). No significant difference between the contralateral and ipsilateral CLA fluorescence was observed (paired t-tests, $p > 0.11$). This indicates that a subset of PFC neurons do indeed make divergent connections to both the DMS and CLA. Furthermore, the lack of significant difference between ipsilateral and contralateral fluorescence suggests that these divergent axons are bilaterally projecting. This histological work and image analysis was carried out in collaboration with undergraduate Niamh Walker.

Thus, the expression of TdTomato observed in the CLA following Cre-dependent expression in DMS-projecting frontal cortical neurons indicates that a subset of neurons in the frontal cortex send divergent input to both downstream structures. This is consistent with the identification of dual laser-responsive neurons identified during in vivo optotagging in Chapter 5.

Appendix B

Appendix B - Developing an optotagging set-up to target projection-defined frontal cortical neurons in vivo

This appendix details the development and optimisation of an in vivo optotagging protocol to target projection-defined frontal neurons

B.1 Introduction

One of the primary goals of neuroscience research is to understand how the brain encodes and utilises information to guide cognition and behaviour. Recent advances in the techniques available for monitoring and manipulating neural activity, particularly in animal models, are opening up novel opportunities to investigate the neural basis of behaviour on a larger scale and with greater specificity than ever before[107;108;150].

These experimental methods operate across a wide range of temporal and spatial scales. Extracellular electrophysiology directly measures neural activity using implanted electrodes to monitor electrical fluctuations in neurons with high temporal resolution. The recent development of high-density silicon probes such as Neuropixels, for instance, is now enabling the activity of hundreds of neurons across different

brain regions to be monitored simultaneously: offering insight into the relationship between behaviour and spiking activity in distributed neural circuits. However, despite the wealth of information electrophysiological approaches can provide about neural activity at both a single-cell and population level, they offer limited insight into the spatial organisation or cell-type identity of recorded neurons.

In contrast, optical strategies have emerged as a means of monitoring neural activity with greater spatial specificity. Calcium imaging using genetically encoded calcium sensors (e.g GCaMP) and wide-field or multi-photon microscopy is now widely utilised to investigate how coordinated patterns of neural activity in spatially restricted areas relate to behaviour. In addition, the use of genetically encoded biosensors - typically delivered using intracranial viral injections - enables specific brain regions or cell populations to be targeted. Such methods can therefore provide insight into the location of specific populations of neurons in space and permit the same neural populations to be recorded longitudinally. However, calcium only represents an indirect measure of neuronal activity, while calcium sensors are characterised by slow temporal dynamics and an inability to detect sub-threshold events. As a result, compared to electrophysiology, optical approaches are limited in their ability to precisely track changes in neural activity, particularly at higher firing rates. Thus, neuroscience remains subject to a paradigmatic challenge reminiscent of Heisenberg's famous uncertainty principle in physics: it's impossible to know precisely where a neuron is and how it's firing simultaneously.

The implications of such methodological limitations are particularly prominent when one considers the inherent complexity of neural circuitry. Neural circuits are comprised of a huge diversity of cell types, which can be defined by their genetic identity, biophysical properties, connectivity, and/or function. While great progress has been made in characterising these different cell types, our understanding of how this heterogeneity contributes to brain function and behaviour remains limited. Addressing these questions requires the integration of multiple experimental approaches capable of examining the functional consequences of multiple properties that define neuronal populations at once.

One such method is photostimulation-assisted identification of neuronal populations[197]: more commonly known as optotagging. Optotagging combines electrophysiological recordings with the expression of opsins in specific groups of neurons with shared genetic and/or anatomical characteristics. The delivery of short, targeted laser pulses is used to drive action potentials in opsin-expressing neurons, such that they can be reliably identified and isolated from non-opsin-expressing cells in electrophysiological recordings. Critically, this method is used to identify and monitor populations of neurons with common characteristics, rather than purposefully manipulating neuronal activity to enhance or disrupt behaviour. Thus, the criticism often levied against optogenetic approaches that the changes in activity they induce are not physiological does not apply in this case. Combined with behaviour, this approach provides an opportunity to probe the role of specific neural pathways in cognition and behaviour.

For example, Ciocchi et al (2015)[113] used optotagging to investigate information routing in ventral hippocampal CA1 (vCA1) neurons in rats. They recorded the activity of vCA1 neurons during an anxiety task, a spatial exploration task, and a goal-directed navigation task, and used projection-specific optotagging to identify which recorded cells projected to the prefrontal cortex, nucleus accumbens, and/or amygdala. The authors observed an over-representation of anxiety-related firing in vCA1 neurons projecting to the prefrontal cortex, while goal-directed firing was prominent in neurons targeting the nucleus accumbens. This suggests that vCA1 neurons selectively route distinct kinds of information to different downstream brain regions. In addition to shedding light on the complex and selective nature of information transfer in the hippocampus, this study highlights the importance of comparing specific neural pathways when investigating the neural encoding of behaviorally relevant information.

For this project, the goal of establishing an optotagging protocol was to isolate frontal cortical neurons projecting specifically to the dorsomedial striatum and claustrum during an in vivo electrophysiological recording. Combining this experimental setup with a decision-making behavioural paradigm would therefore enable us to probe the representation of decision-relevant information in this circuitry. This experimental approach provides a wealth of information about the target population of cells: their

location within the frontal cortex, their downstream projection targets, and their activity in the context of decision-making and reward-motivated behaviour. This appendix will describe the process and results of this period of method development, consider how this approach compares to previously published work, and discuss some of the limitations of in vivo optotagging - both in general and in the context of this project.

B.2 Optimising surgical procedures

To establish an effective and reliable experimental set-up for in vivo optotagging, it was first necessary to optimise the surgical procedures to achieve two key aims: reliable expression of opsin in projection-defined frontal cortical axons and consistent implantation of optic fibers in close apposition to the target downstream structures (Fig B.1A)

This set-up used pAAV-Syn-Chronos-GFP: a high-speed, blue- and green-light activated channelrhodopsin. Channelrhodopsins are a family of light-sensitive proteins, originally derived from photoreceptors in green algae, that permit light-induced control of neuronal activity[108;198]. At the time of release, Chronos was reported to display faster kinetics and greater light sensitivity than any channelrhosopsin that had been produced before[148]. Such properties had great utility in this experiment, as highly fidelity and short latency laser-driven responses were critical for identifying optotagged neurons.

Initially, expression of opsin in frontal neurons was achieved via a single bolus injection of 150nl of Chronos-GFP at coordinates AP: +1.5 mm, ML: +0.6 mm, DV: -1.8 mm. These initial viral injection surgeries were successful in that they resulted in reliable and localised expressed of Chronos-GFP in axons within DMS and CLA. Nonetheless, two adjustments to the injection strategy were made in subsequent cohorts. Firstly, the injection coordinates were moved 0.2mm anterior to capture more anterior regions of frontal cortex (AP: +1.7 mm, ML: +0.6 mm, DV: -1.8 mm). Additionally, the

single bolus injection of 150nl was changed to three smaller injections (50nl) at three different dorsoventral coordinates to achieve opsin-expression - and therefore capture projection defined-cells - across a larger dorsoventral extent of frontal cortex (AP: +1.7 mm, ML: +0.6 mm, DV: [-2.6 -2.2 -1.8] mm). These surgeries resulted in consistent axonal expression of Chronos-GFP in the axons of frontal cortical neurons projecting to DMS and CLA.

The second key aim of these surgical procedures was to consistently implant optic fibers in sufficient proximity to CLA and/or DMS to permit reliable axonal stimulation of opsin-expressing neurons in these areas. The stereotaxic coordinates for fiber implantation were adjusted iteratively until accurate and reliable targeting of these structures was achieved. Another modification was changing the angle of implantation of the DMS optic fiber from vertical to 17°from the vertical to ensure an appropriate amount of space between the two fibers to enable the laser patch cords (Doric) to be connected simultaneously. Visualisation of the fiber tract implanted with these adjusted coordinates confirmed accurate and reliable placement of the fibers above CLA and DMS. The final coordinates used for the implantation of the optic fiber above CLA were AP: +1.3 mm, ML: +2.8 mm, DV: -2.2 mm. The final coordinates used for the DMS optic fiber were AP: +0 mm, ML: +1.2, DV: -2.4 mm and at an implantation angle of 17°. This surgical procedure permitted reliable and accurate targeting of optic fibers above areas of axonal expression of opsin in frontal neurons projecting to DMS and CLA (see Fig B.1C).

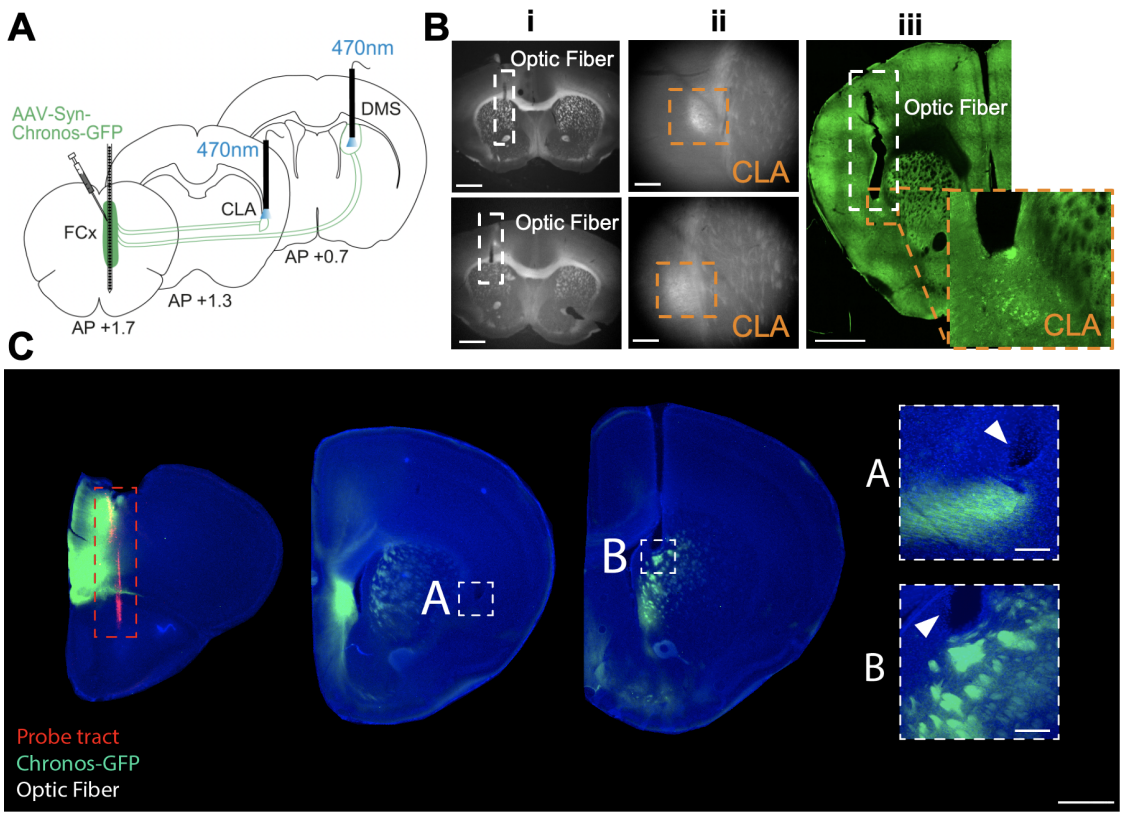


Figure B.1: Refinement of surgical strategy enabled reliable targeting of opsin expression and optic fiber implantation above CLA and DMS A) Schematic of the in vivo optotagging experimental design B) Example histological sections from initial test surgeries i) DMS-fiber implantation test surgeries, scale bar = 1000µm ii) Chronos-GFP injection test surgeries gave rise to visible axonal expression in CLA, scale bar = 160µm iii) A test surgery demonstrating appropriate placement of the CLA fiber, scale bar = 1000µm. CTB was used to retrogradely label CLA in this surgery to verify the fiber was appropriately targeted without needing to wait 3 weeks for opsin expression C) Example histology from one animal used for in vivo optotagging in the final experiment. Post-hoc histology confirmed that the opsin injection (green), optic fibers (white), and Neuropixel probe tract (red) were all correctly targeted, scale bar = 1000µm. A - 10x image of an optic fiber tract above opsin expression in CLA, scale bar = 100µm, B - a 10x image of an optic fiber tract next to opsin expression in DMS, scale bar = 100µm

B.3 Testing the experimental set-up

Once the surgical procedures had been appropriately optimised, it was necessary to test the electrophysiological recording set-up (Fig B.2A) to determine whether the planned recording arrangement was feasible. 'Sham' recordings were conducted

to determine whether the geometry of the recording set-up was realistic using mice implanted with headplates and optic fibers but no intracranial opsin injections (Fig B.2B). Mice were head-fixed in the recording rig and the implanted fibers connected to laser patch cords via silicon sleeves. Several adjustments to the orientation of the recording equipment relative to the implanted optic fibers were necessary to enable movement of the probe without the risk of a collision with the optic fibers.

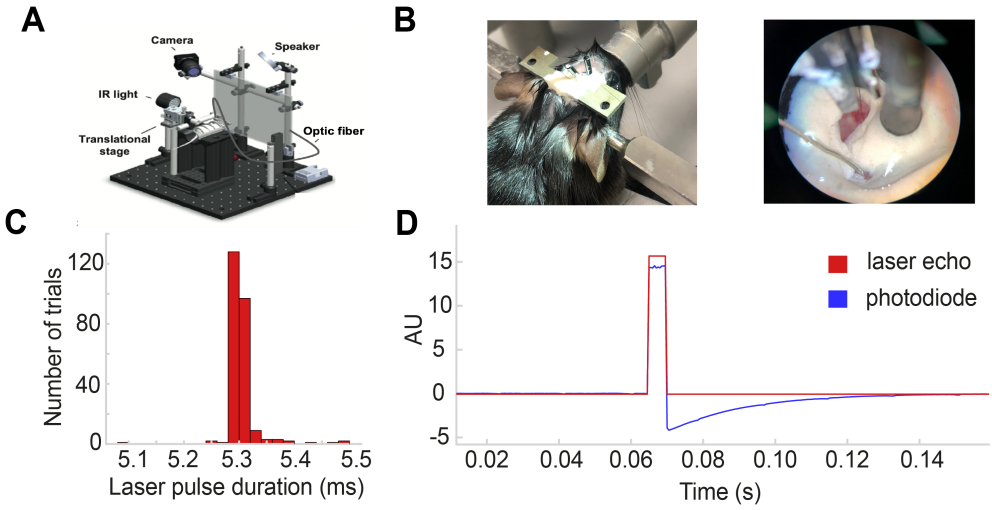


Figure B.2: Testing the experimental set up for in optotagging A) Schematic outlining the rig set-up for in vivo optotagging. The rig set-up is similar to that outlined in Chapter 1 as optotagging will eventually be combined with the head-fixed decision-making task - with the addition of optic patch cords to permit the delivery of laser pulses through chronically implanted optic fibers B) i - A picture outlining the location of the chronically implanted optic fibers above DMS and CLA ii - A microscope image of the recording arrangement during a test recording. The optic fibers connected to patch cords are visible on the right, along with the Neuropixels probe being inserted into an open craniotomy (left) and the grounding electrode (bottom) C) Testing the duration of 5ms laser pulses D) Testing the alignment between laser pulses (via a photodiode, red) and laser echos (blue)

To be confident that laser pulses were driving opsin-expressing neurons to fire directly, rather than as a result of polysynaptic activity, it was also essential to ensure a precise alignment between the laser delivery system and the electrophysiological recording time base. Any drift or lag between these timings would result in a less precise read-out of spike latency following laser onset, and potentially disrupt the ability of this protocol to identify optotagged neurons. It was therefore necessary to test the optotagging protocol itself to establish if it was possible to align the timing of the

electrophysiological recording to the timings of the laser pulses. This was achieved by running the optotagging protocol while directing the laser tip toward a photodiode. The timing of the laser echo trace and the laser pulse picked up by the photodiode were then compared. The mean duration of the 5ms test laser pulses was found to be 5.3ms (Fig B.2C). Comparing laser echo traces with the output from the photodiode further confirmed a tight alignment between these two time-bases.

B.4 Combining optotagging and large-scale electrophysiological recordings

The first full test of the optotagging protocol and recording setup was carried out on an initial cohort of six mice. These mice recieved intracranial injections, optic fiber implantations, and head-plate implantations as described in General methods and materials, Section 2.2.2 - 2.2.3. Animals were left to recover for 3 weeks to permit adequate time for opsin expression before test recordings began. No behavioural training was included at this stage, as the primary purpose of this test cohort was to determine whether it was possible to identify laser-responsive cells from the electrophysiological recording data. A total of seventeen test recording sessions were performed across these six animals.

Once Kilosorted to separate the single and multi-unit activity, the single-unit electrophysiological data was aligned to the laser echo time stamps to examine the effect of laser pulse delivery on both population-level and single neuronal activity. If laser pulses were successfully driving a subset of projection-defined neurons to fire, we would expect to identify a subset of recorded clusters that display a robust increase in firing rate shortly after laser onset.

In all animals (n = 6), alignment of Z-scored trial-averaged population-level activity with laser-pulse onset revealed laser pulse delivery was associated with a rapid excitatory response following laser-onset followed by a sustained period of inhibition across the probe (Fig B.3A-B). The robust inhibition following laser-aligned excitation likely

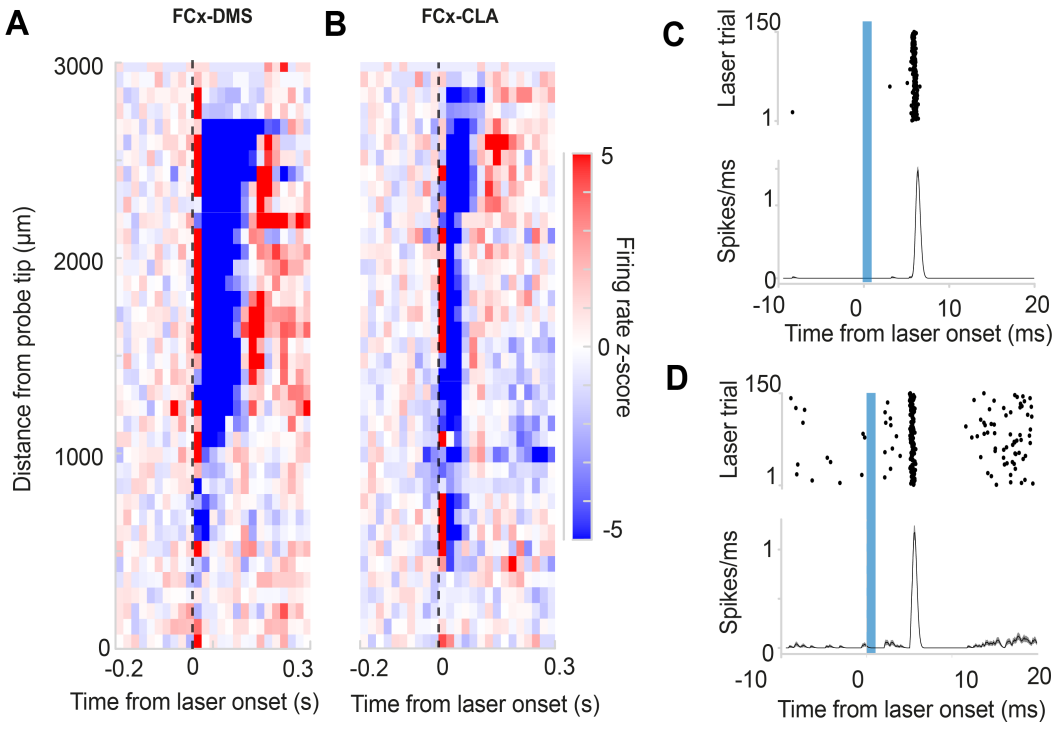


Figure B.3: Laser delivery drove a robust change in firing rate in a subset of recorded neurons A-B) An example of trial-averaged population activity in frontal cortex in response to laser pulse delivery in one recording session. Z-scored firing rate in 80-micron depth bins following DMS (A) and CLA (B) stimulation, 200ms before to 300ms after laser onset C-D) Examples of single units displaying short latency and high fidelity responses to 1ms laser pulses

reflects the activation of local inhibitory circuits in response to laser-evoked activity. The magnitude of the laser-evoked population response at the time of laser-onset, and the subsequent period of inhibition, was variable across mice. This variability was likely the result of differences in viral uptake and subsequently opsin expression, differences in the location of the Neuropixels probe within the same mouse across recording days, and variability in the placement of the optic fibers relative to the area of maximum axonal expression of opsin.

Examining laser-aligned neural responses at a single unit level confirmed that a subset of recorded cells displayed a robust increase in firing rate following laser delivery in all sessions. These cells fired with short latency following laser onset and high fidelity

across laser pulse trials (Fig B.3C -D). Precise numbers of laser-responsive neurons have not been provided here, as the optotagging identification criteria had not been finalised at this time.

Taken together, the laser-driven responses observed at a single-unit and population level across test recordings suggested that this protocol was successfully and consistently driving a subset of projection-defined neurons to fire laser-driven spikes. Such responses could therefore be used to separate dorsomedial striatum and claustrum projecting neurons from a larger population of frontal cortical neurons during a large-scale electrophysiological recording.

However, while this was a promising starting point for the development of this method, it remained a possibility that some seemingly laser-responsive cells were being driven to fire as a result of polysynaptic laser-driven activity, random light scatter hitting off-target opsin expressing neurons following laser onset, or as result of light artefacts. As such, it was necessary to develop a set of quantitative criteria to reliably identify laser-responsive neurons and control for these potential confounds.

B.5 Waveform analysis to validate laser-evoked neural activity

One key question that needed to be addressed in this methodology was how to ensure that any laser-driven activity identified in the recording was physiological, rather than as a result of light artefacts at the time of laser-onset. While Neuropixels are more robust to laser artefacts than previous iterations of extracellular electrophysiological recording strategies[150], it is still possible that direct illumination of the Neuropixel probe would produce an artefact that could be erroneously classified as spiking activity during spikesorting. Such an error could lead a cell to be classified as optotagged when it did not produce physiological laser-driven spikes.

To address this, we developed an analysis pipeline to compare laser-driven and spontaneous spike waveforms. In this analysis, 50 high-quality waveforms were extracted

from each recorded unit in the laser-evoked and spontaneous spiking condition and the Pearson's correlation between the two was calculated. A high correlation between the mean laser-evoked and spontaneous waveforms provides further evidence that stimulus-locked spikes originated from the same unit as spontaneous spikes. The correlation threshold was set at $r > 0.90$ based on the inclusion criteria from other optotagging literature utilising similar recording methods and behavioural approaches[114;186;187]. The median correlation of laser-evoked and spontaneous spikes in the final dataset was far higher than this (DMS-projections: $r = 0.981$; CLA-projections: $r = 0.984$). However, as the main purpose of this criterion was to remove light artefacts and waveforms that did not originate from the same cell, a more coarse threshold - to be used in conjunction with other criteria such as spike latency and reliability - was deemed appropriate in this case. Visualising a subset of these correlations confirmed that this cut-off was capable of separating genuine and artefactual responses (Fig B.4).

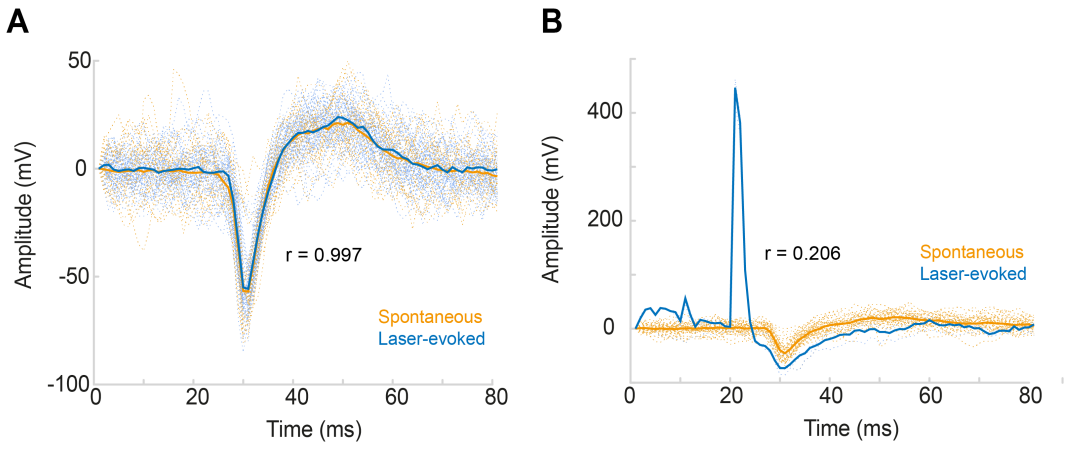


Figure B.4: Example waveform correlation between spontaneous and laser-evoked spikes A) A high correlation between the mean laser-evoked (blue) and spontaneous (orange) spike waveforms B) A low correlation between mean laser-evoked (blue) and spontaneous (orange) spike waveforms. The blue waveform is likely due to a light artefact

Study	Experimental aim	Optotagging criteria
Hirokawa et al (2019)	Encoding of decision-variables in DMS-projecting OFC neurons	SALT Spike latency < 10ms Reliability > 0.4 Waveform Corr. > 0.85
Ciocchi et al (2015)	Information routing in vCA1 neurons projecting to mPFC, Acb and/or Amy during navigational or anxiety-related behaviours	At least three from: Spike latency - (<22ms mPFC, <17ms Acb, <13ms Amy) Reliability > 0.5 Spike jitter < 0.3 ms Collision test Waveform Corr.
Chevee et al (2022)	Sensory responses in S1-projecting claustrum neurons	Spike latency < 7ms Reliability > 0.3 Waveform Corr. > 0.9
Kitanishi et al (2021)	Distribution of information in subiculum neurons to multiple downstream structures (RSC, AV, ITN, MMB) during spatial tasks and sleep	Spike latency < 25ms Reliability > 0.2 Spike jitter < 0.5 ms Collision test Waveform Corr. > 0.9
Bari et al (2019)	Encoding of decision-variables in DMS-projecting mPFC neurons during	Reliability > 0.8 Collision test Somatic tagging Spike latency < 3ms
Zhang et al (2013)	Functional connectivity in entorhinal-hippocampal circuits	CNQX, D-APV
Liu et al (2020)	Neural correlates of visual associative learning in V1-projecting OFC neurons	SALT Waveform Corr. > 0.9 TTX

Table B.1: Projection-specific optotagging criteria in existing literature

B.6 Establishing selection criteria to identify laser-responsive neurons

When implementing an optotagging protocol, it is critical that experimenters can be confident that they are successfully and specifically manipulating their target population of neurons. It is therefore necessary to have a set of metrics for identifying

optotagged cells that minimises the risk of cells being included that have not been directly stimulated by the laser. To this end, we developed a set of criteria based on existing optotagging literature to identify projection-defined frontal cortical neurons in this experiment.

A review of previous studies implementing projection-specific optotagging (Table B.1) highlighted some consistent criteria often used to identify antidromically optotagged cells. The three most common criteria were spike latency, reliability, and waveform correlation.

Spike latency refers to the average time between laser pulse delivery and the first detected spike in recorded cells. In antidromically optotagged neurons, this reflects the conduction time from the stimulation site (axons) to the recording site (the soma). The logic behind imposing a latency cut-off is that neurons directly stimulated by the laser will respond with a shorter latency than those driven to fire as a result of polysynaptic activity. The latency cut-offs used in previous studies varied considerably from 7ms to 25ms. One reason for this variability is that conduction time will vary with axon length, and therefore projections to some downstream brain structures are likely to display longer or shorter spike latencies depending on the distance between the stimulation and the recording site. However, identifying the most appropriate latency cut-off for projections between a given pair of brain regions is challenging, and somewhat circular, if the only information available to judge this is the observed mean latency across the population of putative optotagged neurons. We selected 10ms as our latency cutoff, as this was consistent with existing optotagging studies targeting corticostriatal projections[115] and corticoclaustral projections[186].

Reliability refers to the number of laser pulse trials in which recorded neurons fired a spike within a specific time window post-laser. In the studies above, reliability thresholds varied from $> 20\%$ to 80% of laser pulse trials. While a higher threshold may increase confidence in the fidelity of laser pulse responses for a given cell, it will inevitably reduce the number of cells identified as optotagged: particularly if neurons drift away from the recording electrode during the recording. In contrast,

a lower threshold could result in the erroneous inclusion of cells that are indirectly stimulated, either as a result of polysynaptic input from light-responsive neurons or random light scatter. In this study, reliability was calculated for each unit by dividing the number of trials in which the unit fired a spike within a 10ms window post-laser onset by the total number of laser pulse trials. The reliability threshold was set at >40% of laser pulse trials.

The third most common metric was waveform correlation, addressed in the previous section.

In addition to these three common criteria, optotagged neurons were also required to pass a 'stimulus aligned latency test' (SALT). Initially developed and implemented by Kvistiani et al (2013)[111], SALT is a statistical test to determine whether optogenetic stimulation causes a significant change in spiking activity after laser onset. Specifically, the distribution of first spike latencies relative to the laser pulse, assessed in a 10 ms window after laser onset, was compared to multiple epochs of the same duration in a stimulus-free baseline period. The window used for this comparison was set as 10ms, the same as the spike latency cut-off described above.

Finally, optotagged neurons were required to be deemed 'good' units during Phy curation. A "good" unit was defined as one displaying clear waveforms, a valley-shaped correlogram with fewer than 0.02% refractory period violations, and an attenuation in spike amplitude with increasing distance from the original recording channel. The purpose of these criteria was to ensure that the observed neural responses were not contaminated by multi-unit activity. It would be challenging to draw conclusions about the activity and function of a specific population of projection-defined neurons if their activity were combined with other, non-projecting neurons.

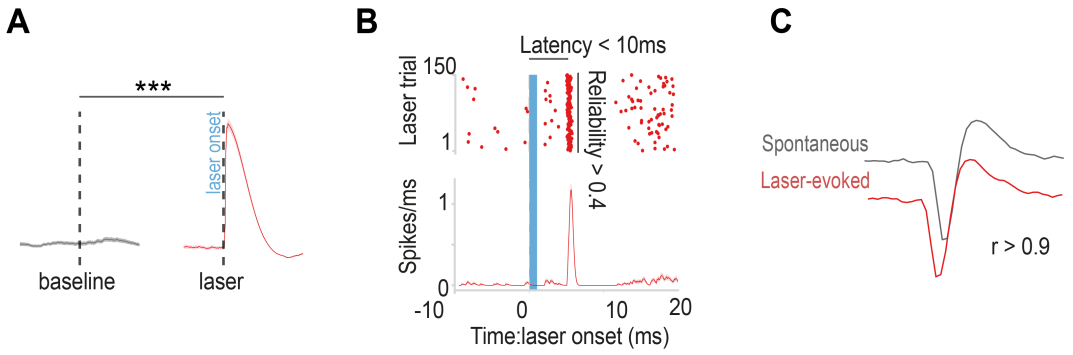


Figure B.5: In vivo optotagging criteria for identifying CLA- and DMS-projecting frontal cortical neurons neurons. To be deemed 'optotagged', laser-responsive neurons had to display A) A significant SALT (stimulus-aligned latency test[111]) statistic, indicating a significant difference in firing rate relative to baseline in a 10ms window post laser-onset B) A mean first-spike latency <10ms and reliability of >40% C) A strong correlation between spontaneous and laser evoked spikes (r > 0.9)

Overall, guided by existing literature, we developed a strict set of quantitative criteria to identify optotagged neurons based on the latency, fidelity, and quality of their laser-driven responses. In addition, these criteria served to significantly reduce the risk that indirect, artefactual, or multi-unit responses were erroneously included in the dataset. For the final numbers of optotagged neurons identified using these criteria, please see Chapter 5 Section 5.2.1 (In vivo optotagging identified populations of DMS and CLA-projecting frontal neurons).

B.7 Collision testing

One general limitation of optotagging is that it depends wholly on the timing of laser-driven spikes to identify projection-defined cells. This typically involves imposing a latency cut-off to separate the neurons displaying reliable laser-driven responses from those driven to fire indirectly by polysynaptic inputs, as noted above. However, depending on the cut-off used, there remains a possibility that some indirectly stimulated cells may fire with a sufficiently short latency to pass the selected threshold. As such, some studies (e.g Ciocchi et al, 2015; Kitanishi et al, 2021; Bari et al, 2019) implement an additional criterion known as a 'collision test'.

Collision testing is the gold-standard approach for identifying projection-defined neurons using antidromic optotagging. It is based on the principle that two spikes travelling in opposite directions along the same axon will collide and dissipate. This behaviour has been demonstrated in neurons[199]. Antidromically stimulating the axon of a neuron will send a spike propagating backwards along the axon, where it will be detected at the soma a few milliseconds later (Fig B.5A - top). However, if an antidromic spike is fired just after a spontaneous spike, these two spikes will collide. As a result, no antidromic spike will be detected at the soma (Fig B.5A - middle). In contrast, no collisions will occur if the recorded neuron is being driven to fire indirectly via polysnaptic inputs from stimulated neurons (Fig B.5A - bottom). Thus, the presence of collisions when spontaneous and laser-evoked spikes occur close together in time provides conclusive evidence that a cell is being directly stimulated by the laser.

We sought to investigate the potential utility of collision testing in this experimental setup by investigating the change in neural response efficacy when laser-responsive neurons emitted spontaneous spikes within a critical time window before laser delivery. The critical window was based on the mean spike latency after laser onset, as this indicates how long it takes for an antidromic spike to travel from the axon to the soma in each neuron. Based on the collision test principle, we hypothesised that neurons producing spontaneous spikes in this window would display either increased spike latency or decreased fidelity. A Chi-squared test was employed to evaluate the difference in reliability between 'collision' and 'non-collision' trials. For example, Figure B.5B outlines an optotagged unit that displays a clear reduction in laser response fidelity when spontaneous spikes occur within the critical window before laser onset.

However, we met with some practical challenges in the implementation of this test. Initially, the optotagging protocol consisted of 50 laser pulse trials per stimulation block. In these recordings, the frequency with which neurons fired spontaneous spikes within the critical window was insufficient to perform the test. To address this, the number of laser pulses in each optotagging block was increased to 150 and then 500 pulses. While these adjustments resulted in a higher frequency of neurons firing

spontaneous spikes within this window, it became evident that this cut-off would bias the sampling towards more highly active cells by excluding cells with a low baseline firing rate that failed to produce a sufficient number of spikes within the defined collision window.

Thus, despite recognising the collision test as the gold standard for confirming direct antidromic stimulation, the practical limitations encountered in this case — largely stemming from the need for a substantial number of collision trials and the bias towards higher activity cells — suggest its limitations outweigh its benefits in this experimental context.

In the future, it may be possible to address these issues using an integrated closed-loop system for online collision testing. In this kind of setup, it is possible to identify laser-responsive neurons in real-time and trigger antidromic spikes to fire automatically following a spontaneous spike. However, the technical requirements for such a system - which would include specialised equipment and software (e.g., Multi-Linc, Mitani et al., 2022[200]) - were beyond the scope of the current project. As such, this represents a refinement for future work.

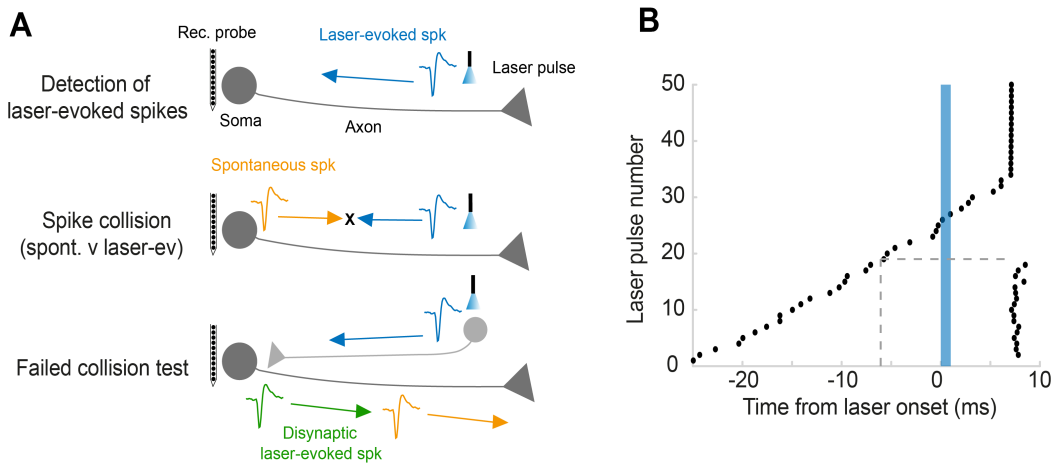


Figure B.6: Collision testing A) A schematic outlining the principle of a collision test in antidromic optotagging. Top - axonal stimulation of an opsin expressing neuron will generate an antidromic spike that will be detected at the recording probe a few milliseconds later. Middle - If a neuron produces a spontaneous somatic spike just before a laser-driven antidromic spike, the two will collide and cancel out. As such, no antidromic spike will be detected at the soma. Bottom - If a neuron receives indirect laser-stimulation via another neuron, no collisions will occur between disynaptic laser-evoked spikes and spontaneous spikes. B) An example neuron passing the collision test. When the neuron generates spontaneous spikes in the critical window before laser onset, there is a reduction in the fidelity of laser-evoked spikes. The window for assessing spike collisions is defined by the mean laser-evoked spike latency, as this is the time taken to travel from the stimulation site to the soma

B.8 Discussion

The purpose of this period of method development was to establish an optotagging protocol that enabled projection-defined neurons in the frontal cortex to be isolated from a larger population of neurons during an in vivo electrophysiological recording. This was achieved by expressing an opsin in the frontal cortex and antidromically stimulating the axons of DMS- and CLA-projecting frontal cortical neurons via laser pulses delivered through optic fibers implanted above these downstream structures.

This period of optimisation led to the development of an experiment protocol that combined simultaneous large-scale electrophysiological recordings in frontal cortex and antidromic optotagging of DMS- and CLA-projecting frontal cortical neurons.

Examining laser-aligned population activity revealed laser onset drove rapid excitation across the recording probe, suggesting that laser delivery successfully drove an increase in firing rate in a subset of frontal cortical neurons. This was subsequently confirmed by examining single-unit laser-aligned responses, which identified neurons displaying high fidelity and short latency increases in firing rate following laser-onset. We then developed a rigorous set of criteria based on existing optotagging literature to separate laser-responsive neurons from those driven to fire by polysynaptic activity, light artefacts erroneously classified as spikes, or as a result of multi-unit activity misattributed to individual cells.

Optotagging leverages the high temporal resolution of electrophysiology to provide precise insight into spiking activity, while offering some of the cell-type specificity observed in optical and optogenetic approaches. As a result, a successful optotagging protocol represents a valuable tool for physiologists investigating the neural basis of behaviour, enabling specific populations of neurons across the brain to be identified and monitored in behaving animals. In this project, this method will be utilised to examine the activity of DMS- and CLA-projecting frontal neurons in the context of decision-making behaviour.

B.8.1 Limitations

One of the challenges of an experimental set-up that seeks to target distinct populations of cells is the inevitable trade-off between specificity and throughput. While the dorsomedial striatum and claustrum both receive dense input from frontal cortical regions, they still only represent a small portion of the overall output of frontal cortex. As such, there is always a concern that reasonable steps to improve specificity and confidence in the identity of optotagged cells could result in a significant reduction in data throughput if only a small subset of the target population is captured.

For example, one means of increasing confidence in the identity of optotagged neurons is the use of conditional expression systems such as Cre-LoxP[201]. Injections of a floxed opsin in the frontal cortex could be combined with retro-Cre injections in DMS and

CLA to ensure that opsin expression was restricted to these projection pathways. However, such a system would inevitably reduce throughput in this experiment as not all cells will successfully take up and co-express both viruses. While the non-conditional expression of Chronos-GFP increases the likelihood of scattered photons hitting off-target opsin-expressing neurons, efforts to counter-balance this were made by introducing the selection criteria described above to ensure only reliably laser-responsive cells driven to fire with short latency were deemed to be optotagged. Further evidence in favour of the non-conditional expression being suitably specific was provided by the observation that animals in which the fibers were incorrectly targeted did not produce any optotagged cells that passed the necessary criteria for inclusion. This suggests that the placement of optic fibers in close proximity to DMS and CLA was required for the identification of optotagged neurons, and did not result in off-target stimulation.

Another limitation of this method is that we cannot wholly exclude the possibility that some of the putative projection-defined neurons were non-projecting but indirectly stimulated due to the synchronous activation of optotagged neurons rapidly exciting a postsynaptic, non-projecting neuron. However, the criteria used to identify optotagged neurons were similar to standards in the field[111]. Thus, while we cannot definitely rule out the possibility that our optotagging strategy misidentified a subset of non-projecting neurons, this protocol remains comparable to other methods implemented in the field in recent years.

One additional control that would increase confidence that only directly laser-responsive neurons are included in the dataset is the application NMDA or AMPA receptor antagonists such as NBQX or D-AP5. These antagonists inhibit excitatory synaptic transmission but would not prevent neurons from generating laser-evoked action potentials. As such, infusions of NBQX/D-AP5 would eliminate any polysynaptic laser-evoked activity while leaving direct laser responses intact. This is a common strategy for confirming the direct stimulation of neurons during in vitro optotagging, as it is reasonably straightforward to wash slices in these antagonists. However, in the context of this experiment, the depth with which these recordings were carried out would

make it challenging to deliver infusions of NBQX/D-AP5 across all the recorded regions of the frontal cortex. Thus, while the rationale for this approach is clear, the practical challenges make this particular control unrealistic for this experiment.

B.8.2 Summary

Overall, this appendix describes how the in vivo optotagging approach used in Chapter 5 was developed and highlights its consistency with previously published methods of this kind. This represents a valuable tool for probing the neural basis of behaviour in distinct neuronal populations.

Acknowledgments

They say it takes a village to raise a child, and the same could be said of completing a DPhil. This journey would not have been possible without the support and contributions of so many incredible people.

Firstly, to Armin: I don't think any of us fully comprehend how much you do to keep the lab running. Thank you for working tirelessly to create the environment and resources that make this work possible, and for believing in me — and reminding me not to stress too much — when I was convinced it was all going wrong. Your guidance and supervision have been invaluable. And yes, I can start working on the paper now.

None of this work would have been possible without my wonderful colleagues, collaborators, and friends: Antara and Matthias. Antara, we've been in this together since the very beginning, and I couldn't have asked for a better lab mate and desk buddy to share these experiences with — from the highs of beautiful spikes to the lows of code that won't run or mice that won't learn. Matthias, I don't think I'm exaggerating when I say this document wouldn't exist without your patience, generosity with your time and knowledge, and consistently insightful contributions and advice. You're going to make an excellent PI, and any student would be lucky to have you as a mentor.

To the rest of the Lak Lab — Blake, Chiara, Jess, Lauren, Marko, Michael, Olena, Sam, Sandra — thank you for all the good times, endless support, and for laughing at my jokes and pixel art figures in lab meetings. A special thank you to Peter, the unsung hero behind everything that works in the lab, for your friendship and willingness to help a PhD student in need years after leaving academia. To the various members, past and present, of the Packer and Butt Labs —Andrew, David, Gabi, Hamish, Huriye, Ivan, Jimmy, Rob, Sarah, Thijs — thanks for making the DPAG Penthouse the place to be and for making me feel at home from the very start. And to Adam, Kerry, and Simon — thank you for the important roles each of

you has played in my time in Oxford. I can identify so many moments where your mentorship, guidance, and decision to take a chance on a keen aspiring neuroscientist made a significant difference to the course of my academic journey.

The past few years wouldn't have been the same without the light and love of my amazing friends. Owen, Tanya, and Barley: thank you for keeping me sane, for always being there for me, and for giving me some of my favourite memories of Oxford. I'm so grateful to you all. To my favourite members of the Oriel Shadow SCR — Marta, Tishan, Cécile, John, Val, and Azmi — it's been such a joy getting to know you all this year. You've made the final year of my DPhil so much more fun than I was expecting it to be. A special thanks also to Lucy: from morning runs and surf trips to hours spent working in my living room and ranting about science, you've added so much to my life since we met. Your friendship is truly a gift.

I owe a huge amount to my family, both in New Zealand and the UK: you've moved mountains to help me follow my dreams. Thanks especially to my grandparents — Barbara, Jean, Steve, Anna, and Neville — and my parents and brother — Jo, Anton, and Reuben — for all your love and support over the years. Thanks also to the Greenoughs — Kate, John, and Meg — for welcoming me into your family and for showing me so much care and generosity for all the time I've known you.

And finally, to Will. You've been the best partner, teammate, confidant, editor-in-chief, and cheerleader anyone could ask for. This process hasn't always been easy — you've had to deal with more than your fair share of tears, stress, and frustration — but your love, support, humour, pragmatism, and unwavering faith in me have made all the difference. I wouldn't be the person I am today without you.

Bibliography

[1] John Von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ, USA, 1944.

[2] Leonard J. Savage. *The Foundations of Statistics*. Wiley Publications in Statistics, 1954.

[3] Daniel Kahneman and Amos Tversky. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263–291, 1979. Publisher: [Wiley, Econometric Society].

[4] Daniel Kahneman and Amos Tversky. Choices, values, and frames. *American Psychologist*, 39(4):341–350, 1984. Place: US Publisher: American Psychological Association.

[5] John M. Harlow. Recovery from the passage of an iron bar through the head. *History of Psychiatry*, 4(14):274–281, June 1993. Publisher: SAGE Publications Ltd.

[6] Wilder Penfield and Joseph Evans. The frontal lobe in man: a clinical study of maximu removals. *Brain*, 58(1):115–133, March 1935.

[7] Brenda Milner. Effects of Different Brain Lesions on Card Sorting: The Role of the Frontal Lobes. *Archives of Neurology*, 9(1):90, July 1963.

[8] Paul J. Eslinger and Antonio R. Damasio. Severe disturbance of higher cognition after bilateral frontal lobe ablation. *Neurology*, 35(12):1731–1731, December 1985. Publisher: Wolters Kluwer.

[9] Tim Shallice and Paul Burgess. Deficits in strategy application following frontal lobe damage in man. *Brain*, 114(2):727–741, April 1991.

[10] Antoine Bechara, Antonio R. Damasio, Hanna Damasio, and Steven W. Anderson. Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, 50(1):7–15, April 1994.

[11] Facundo Manes, Barbara Sahakian, Luke Clark, Robert Rogers, Nagui Antoun, Mike Aitken, and Trevor Robbins. Decision-making processes following damage to the prefrontal cortex. *Brain*, 125(3):624–639, March 2002.

[12] Camillo Padoa-Schioppa and John A. Assad. Neurons in the orbitofrontal cortex encode economic value. *Nature*, 441(7090):223–226, May 2006. Number: 7090 Publisher: Nature Publishing Group.

[13] Léon Tremblay and Wolfram Schultz. Relative reward preference in primate orbitofrontal cortex. *Nature*, 398(6729):704–708, April 1999. Publisher: Nature Publishing Group.

[14] Jonathan D. Wallis and Earl K. Miller. Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. *European Journal of Neuroscience*, 18(7):2069–2081, 2003. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1460-9568.2003.02922.x.

[15] Adam Kepecs, Naoshige Uchida, Hatim A. Zariwala, and Zachary F. Mainen. Neural correlates, computation and behavioural impact of decision confidence. *Nature*, 455(7210):227–231, September 2008.

[16] Johannes Passecker, Nace Mikus, Hugo Malagon-Vina, Philip Anner, Jordane Dimidschstein, Gordon Fishell, Georg Dorffner, and Thomas Klausberger. Activity of Prefrontal Neurons Predict Future Choices during Gambling. *Neuron*, 101(1):152–164.e7, January 2019. Publisher: Elsevier.

[17] Bilal A. Bari, Cooper D. Grossman, Emily E. Lubin, Adithya E. Rajagopalan, Jianna I. Cressy, and Jeremiah Y. Cohen. Stable Representations of Decision Variables for Flexible Behavior. *Neuron*, 103(5):922–933.e7, September 2019.

[18] D. O. Hebb. Intelligence in Man after Large Removals of Cerebral Tissue: Report of Four Left Frontal Lobe Cases. *The Journal of General Psychology*, 21(1):73–87, July 1939. Publisher: Routledge _eprint: https://doi.org/10.1080/00221309.1939.9710587.

[19] Asenath Petrie. *Personality and the frontal lobes: An investigation of the psychological effects of different types of leucotomy.* Personality and the frontal lobes: An investigation of the psychological effects of different types of leucotomy. Routledge & Kegan Paul, London, England, 1952. Pages: xii, 188.

[20] H. L. Teuber, W. S. Battersby, and M. B. Bender. Performance of complex visual tasks after cerebral lesions. *The Journal of Nervous and Mental Disease*, 114(5):413–429, November 1951.

[21] Sara M. Szczepanski and Robert T. Knight. Insights into Human Behavior from Lesions to the Prefrontal Cortex. *Neuron*, 83(5):1002–1018, September 2014.

[22] Timothy Shallice. Specific impairments of planning. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 298(1089):199–209, June 1982.

[23] Adrian M. Owen, John J. Downes, Barbara J. Sahakian, Charles E. Polkey, and Trevor W. Robbins. Planning and spatial working memory following frontal lobe lesions in man. *Neuropsychologia*, 28(10):1021–1034, January 1990.

[24] A. R. Luria. Frontal Lobe Syndromes. In P. J. Vinken and G. W. Bruyn, editors, *Handbook of Clinical Neurology*, pages 2–725. North Holland, 1969.

[25] Michael Petrides and Brenda Milner. Deficits on subject-ordered tasks after frontal- and temporal-lobe lesions in man. *Neuropsychologia*, 20(3):249–262, January 1982.

[26] Antonio R. Damasio, Daniel Tranel, and Hanna C. Damasio. Somatic markers and the guidance of behavior: Theory and preliminary testing. In *Frontal lobe function and dysfunction*, pages 217–229. Oxford University Press, New York, NY, US, 1991.

[27] Antoine Bechara, Daniel Tranel, Hanna Damasio, and Antonio R. Damasio. Failure to Respond Autonomically to Anticipated Future Outcomes Following Damage to Prefrontal Cortex. *Cerebral Cortex*, 6(2):215–225, 1996.

[28] Antoine Bechara, Hanna Damasio, Daniel Tranel, and Antonio R. Damasio. Deciding Advantageously Before Knowing the Advantageous Strategy. *Science*, 275(5304):1293–1295, February 1997. Publisher: American Association for the Advancement of Science.

[29] Antoine Bechara, Hanna Damasio, Daniel Tranel, and Steven W. Anderson. Dissociation Of Working Memory from Decision Making within the Human Prefrontal Cortex. *Journal of Neuroscience*, 18(1):428–437, January 1998. Publisher: Society for Neuroscience Section: ARTICLE.

[30] Robert D. Rogers, Adrian M. Owen, Hugh C. Middleton, Emma J. Williams, John D. Pickard, Barbara J. Sahakian, and Trevor W. Robbins. Choosing between Small, Likely Rewards and Large, Unlikely Rewards Activates Inferior and Orbital Prefrontal Cortex. *Journal of Neuroscience*, 19(20):9029–9038, October 1999. Publisher: Society for Neuroscience Section: ARTICLE.

[31] R. D. Rogers, B. J. Everitt, A. Baldacchino, A. J. Blackshaw, R. Swainson, K. Wynne, N. B. Baker, J. Hunter, T. Carthy, E. Booker, M. London, J. F. W. Deakin, B. J. Sahakian, and T. W. Robbins. Dissociable Deficits in the Decision-Making Cognition of Chronic Amphetamine Abusers, Opiate Abusers, Patients with Focal Damage to Prefrontal Cortex, and Tryptophan-Depleted Normal Volunteers: Evidence for Monoaminergic Mechanisms. *Neuropsychopharmacology*, 20(4):322–339, April 1999. Publisher: Nature Publishing Group.

[32] Alan I. Leshner. Addiction Is a Brain Disease, and It Matters. *Science*, 278(5335):45–47, October 1997. Publisher: American Association for the Advancement of Science.

[33] Volkow Nora D., Koob George F., and McLellan A. Thomas. Neurobiologic Advances from the Brain Disease Model of Addiction. *New England Journal of Medicine*, 374(4):363–371, 2016. Publisher: Massachusetts Medical Society _eprint: https://www.nejm.org/doi/pdf/10.1056/NEJMra1511480.

[34] Paula Banca, Martin D. Vestergaard, Vladan Rankov, Kwangyeol Baek, Simon Mitchell, Tatyana Lapa, Miguel Castelo-Branco, and Valerie Voon. Evidence Accumulation in Obsessive-Compulsive Disorder: the Role of Uncertainty and Monetary Reward on Perceptual Decision-Making Thresholds. *Neuropsychopharmacology*, 40(5):1192–1202, April 2015. Publisher: Nature Publishing Group.

[35] Philipp Sterzer, Martin Voss, Florian Schlagenhauf, and Andreas Heinz. Decision-making in schizophrenia: A predictive-coding perspective. *NeuroImage*, 190:133–143, April 2019.

[36] Neir Eshel and Jonathan P. Roiser. Reward and Punishment Processing in Depression. *Biological Psychiatry*, 68(2):118–124, July 2010.

[37] Robert Hester, Veronica Dixon, and Hugh Garavan. A consistent attentional bias for drug-related material in active cocaine users across word and picture versions of the emotional Stroop task. *Drug and Alcohol Dependence*, 81(3):251–257, February 2006.

[38] Anna Zilverstand, Anna S. Huang, Nelly Alia-Klein, and Rita Z. Goldstein. Neuroimaging Impaired Response Inhibition and Salience Attribution in Human Drug Addiction. A systematic review. *Neuron*, 98(5):886–903, June 2018.

[39] M.A. Parvaz, A.B. Konova, G.H. Proudfit, J.P. Dunning, P. Malaker, S.J. Moeller, T. Maloney, N. Alia-Klein, and R.Z. Goldstein. Impaired neural response to negative prediction errors in cocaine addiction. *Journal of Neuroscience*, 35(5):1872–1879, 2015.

[40] Muhammad A. Parvaz, Scott J. Moeller, and Rita Z. Goldstein. Incubation of Cue-Induced Craving in Adults Addicted to Cocaine Measured by Electroencephalography. *JAMA psychiatry*, 73(11):1127–1134, November 2016.

[41] Rita Z. Goldstein and Nora D. Volkow. Drug Addiction and Its Underlying Neurobiological Basis: Neuroimaging Evidence for the Involvement of the Frontal Cortex. *American Journal of Psychiatry*, 159(10):1642–1652, October 2002. Publisher: American Psychiatric Publishing.

[42] Rita Z. Goldstein and Nora D. Volkow. Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications. *Nature reviews. Neuroscience*, 12(11):652–669, October 2011.

[43] Aaron T. Beck. The Evolution of the Cognitive Model of Depression and Its Neurobiological Correlates. *American Journal of Psychiatry*, 165(8):969–977, August 2008. Publisher: American Psychiatric Publishing.

[44] James W. Murrough, Brian Iacoviello, Alexander Neumeister, Dennis S. Charney, and Dan V. Iosifescu. Cognitive dysfunction in depression: Neurocircuitry and new therapeutic strategies. *Neurobiology of Learning and Memory*, 96(4):553–563, November 2011.

[45] L. D. Selemon and P. S. Goldman-Rakic. Common cortical and subcortical targets of the dorsolateral prefrontal and posterior parietal cortices in the rhesus monkey: evidence for a distributed neural network subserving spatially guided behavior. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 8(11):4049–4068, November 1988.

[46] S. T. Carmichael and J. L. Price. Architectonic subdivision of the orbital and medial prefrontal cortex in the macaque monkey. *Journal of Comparative Neurology*, 346(3):366–402, 1994. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cne.903460305.

[47] D. Öngür and J.L. Price. The Organization of Networks within the Orbital and Medial Prefrontal Cortex of Rats, Monkeys and Humans. *Cerebral Cortex*, 10(3):206–219, March 2000.

[48] Le Gao, Sang Liu, Lingfeng Gou, Yachuang Hu, Yanhe Liu, Li Deng, Danyi Ma, Haifang Wang, Qiaoqiao Yang, Zhaoqin Chen, Dechen Liu, Shou Qiu, Xiaofei Wang, Danying Wang, Xinran Wang, Biyu Ren, Qingxu Liu, Tianzhi Chen, Xiaoxue Shi, Haishan Yao, Chun Xu, Chengyu T. Li, Yangang Sun, Anan Li, Qingming Luo, Hui Gong, Ninglong Xu, and Jun Yan. Single-neuron projectome of mouse prefrontal cortex. *Nature Neuroscience*, 25(4):515–529, April 2022. Number: 4 Publisher: Nature Publishing Group.

[49] Julie A. Harris, Stefan Mihalas, Karla E. Hirokawa, Jennifer D. Whitesell, Hannah Choi, Amy Bernard, Phillip Bohn, Shiella Caldejon, Linzy Casal, Andrew Cho, Aaron Feiner, David Feng, Nathalie Gaudreault, Charles R. Gerfen, Nile Graddis, Peter A. Groblewski, Alex M. Henry, Anh Ho, Robert Howard, Joseph E. Knox, Leonard Kuan, Xiuli Kuang, Jerome Lecoq, Phil Lesnar, Yaoyao Li, Jennifer Luviano, Stephen McConoughey, Marty T. Mortrud, Maitham Naeemi, Lydia Ng, Seung Wook Oh, Benjamin Ouellette, Elise Shen, Staci A. Sorensen, Wayne Wakeman, Quanxin Wang, Yun Wang, Ali Williford, John W. Phillips, Allan R. Jones, Christof Koch, and Hongkui Zeng. Hierarchical organization of cortical and thalamic connectivity. *Nature*, 575(7781):195–202, November 2019. Publisher: Nature Publishing Group.

[50] Mark J. Buckley, Farshad A. Mansouri, Hassan Hoda, Majid Mahboubi, Philip G. F. Browning, Sze C. Kwok, Adam Phillips, and Keiji Tanaka. Dissociable Components of Rule-Guided Behavior Depend on Distinct Medial and Prefrontal Regions. *Science*, 325(5936):52–58, July 2009. Publisher: American Association for the Advancement of Science.

[51] Ami Tsuchida and Lesley K. Fellows. Lesion evidence that two distinct regions within prefrontal cortex are critical for n-back performance in humans. *Journal of Cognitive Neuroscience*, 21(12):2263–2275, December 2009.

[52] L. L. Chao and R. T. Knight. Contribution of human prefrontal cortex to delay performance. *Journal of Cognitive Neuroscience*, 10(2):167–177, March 1998.

[53] Brian Knutson, Jonathan Taylor, Matthew Kaufman, Richard Peterson, and Gary Glover. Distributed Neural Representation of Expected Value. *Journal of Neuroscience*, 25(19):4806–4812, May 2005. Publisher: Society for Neuroscience Section: Behavioral/Systems/Cognitive.

[54] Joseph W. Kable and Paul W. Glimcher. The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience*, 10(12):1625–1633, December 2007. Publisher: Nature Publishing Group.

[55] Dino J Levy and Paul W Glimcher. The root of all value: a neural common currency for choice. *Current Opinion in Neurobiology*, 22(6):1027–1038, December 2012.

[56] Nathaniel D. Daw, John P. O'Doherty, Peter Dayan, Ben Seymour, and Raymond J. Dolan. Cortical substrates for exploratory decisions in humans. *Nature*, 441(7095):876–879, June 2006.

[57] Jan Gläscher, Alan N. Hampton, and John P. O'Doherty. Determining a Role for Ventromedial Prefrontal Cortex in Encoding Action-Based Value Signals During Reward-Related Decision Making. *Cerebral Cortex*, 19(2):483–495, February 2009.

[58] Antoine Bechara, Daniel Tranel, and Hanna Damasio. Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions. *Brain*, 123(11):2189–2202, November 2000.

[59] Nathalie Camille, Cathryn A. Griffiths, Khoi Vo, Lesley K. Fellows, and Joseph W. Kable. Ventromedial Frontal Lobe Damage Disrupts Value Maximization in Humans. *The Journal of Neuroscience*, 31(20):7527–7532, May 2011.

[60] L. K. Fellows and M. J. Farah. The Role of Ventromedial Prefrontal Cortex in Decision Making: Judgment under Uncertainty or Judgment Per Se? *Cerebral Cortex*, 17(11):2669–2674, November 2007.

[61] Matthew F. S. Rushworth, MaryAnn P. Noonan, Erie D. Boorman, Mark E. Walton, and Timothy E. Behrens. Frontal Cortex and Reward-Guided Learning and Decision-Making. *Neuron*, 70(6):1054–1069, June 2011.

[62] Todd A. Hare, John O'Doherty, Colin F. Camerer, Wolfram Schultz, and Antonio Rangel. Dissociating the Role of the Orbitofrontal Cortex and the Striatum in the Computation of Goal Values and Prediction Errors. *Journal of Neuroscience*, 28(22):5623–5630, May 2008. Publisher: Society for Neuroscience Section: Articles.

[63] Mark E. Walton, Timothy E. J. Behrens, Mark J. Buckley, Peter H. Rudebeck, and Matthew F. S. Rushworth. Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. *Neuron*, 65(6):927–939, March 2010.

[64] Cameron S. Carter, Todd S. Braver, Deanna M. Barch, Matthew M. Botvinick, Douglas Noll, and Jonathan D. Cohen. Anterior Cingulate Cortex, Error Detection, and the Online Monitoring of Performance. *Science*, 280(5364):747–749, May 1998. Publisher: American Association for the Advancement of Science.

[65] John G. Kerns, Jonathan D. Cohen, Angus W. MacDonald, Raymond Y. Cho, V. Andrew Stenger, and Cameron S. Carter. Anterior Cingulate Conflict Monitoring and Adjustments in Control. *Science*, 303(5660):1023–1026, February 2004. Publisher: American Association for the Advancement of Science.

[66] Mark E. Walton, David M. Bannerman, and Matthew F. S. Rushworth. The Role of Rat Medial Frontal Cortex in Effort-Based Decision Making. *Journal of Neuroscience*, 22(24):10996–11003, December 2002. Publisher: Society for Neuroscience Section: ARTICLE.

[67] Mark E. Walton, David M. Bannerman, Karin Alterescu, and Matthew F. S. Rushworth. Functional Specialization within Medial Frontal Cortex of the Anterior Cingulate for Evaluating Effort-Related Decisions. *Journal of Neuroscience*, 23(16):6475–6479, July 2003. Publisher: Society for Neuroscience Section: BRIEF COMMUNICATIONS.

[68] Peter H. Rudebeck, Mark E. Walton, Angharad N. Smyth, David M. Bannerman, and Matthew F. S. Rushworth. Separate neural pathways process different decision costs. *Nature Neuroscience*, 9(9):1161–1168, September 2006. Number: 9 Publisher: Nature Publishing Group.

[69] Steven W. Kennerley, Mark E. Walton, Timothy E. J. Behrens, Mark J. Buckley, and Matthew F. S. Rushworth. Optimal decision making and the anterior cingulate cortex. *Nature Neuroscience*, 9(7):940–947, July 2006. Publisher: Nature Publishing Group.

[70] Abigail A. Marsh, Karina S. Blair, Meena Vythilingam, Sarah Busis, and R. J. R. Blair. Response options and expectations of reward in decision-making: The differential roles of dorsal and rostral anterior cingulate cortex. *NeuroImage*, 35(2):979–988, April 2007.

[71] Earl K. Miller and Jonathan D. Cohen. An Integrative Theory of Prefrontal Cortex Function. *Annual Review of Neuroscience*, 24(1):167–202, 2001. _eprint: https://doi.org/10.1146/annurev.neuro.24.1.167.

[72] P.S Goldman-Rakic. Cellular basis of working memory. *Neuron*, 14(3):477–485, March 1995.

[73] Alexander G. E. Fuster, J. M. Neuron Activity Related to Short-Term Memory. *Science*, 173(3997):652–654, March 1971.

[74] S. Funahashi, C. J. Bruce, and P. S. Goldman-Rakic. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology*, 61(2):331–349, February 1989. Publisher: American Physiological Society.

[75] H. Niki and M. Watanabe. Prefrontal and cingulate unit activity during timing behavior in the monkey. *Brain Research*, 171(2):213–224, 1979.

[76] Masataka Watanabe. Reward expectancy in primate prefrental neurons. *Nature*, 382(6592):629–632, August 1996. Number: 6592 Publisher: Nature Publishing Group.

[77] Matthew I. Leon and Michael N. Shadlen. Effect of Expected Reward Magnitude on the Response of Neurons in the Dorsolateral Prefrontal Cortex of the Macaque. *Neuron*, 24(2):415–425, October 1999. Publisher: Elsevier.

[78] Matthew R. Roesch and Carl R. Olson. Impact of Expected Reward on Neuronal Activity in Prefrontal Cortex, Frontal and Supplementary Eye Fields and Premotor Cortex. *Journal of Neurophysiology*, 90(3):1766–1789, September 2003. Publisher: American Physiological Society.

[79] Matthew R. Roesch and Carl R. Olson. Neuronal Activity Related to Reward Value and Motivation in Primate Frontal Cortex. *Science*, 304(5668):307–310, April 2004. Publisher: American Association for the Advancement of Science.

[80] Jonathan D. Wallis, Kathleen C. Anderson, and Earl K. Miller. Single neurons in prefrontal cortex encode abstract rules. *Nature*, 411(6840):953–956, June 2001. Number: 6840 Publisher: Nature Publishing Group.

[81] Geoffrey Schoenbaum, Andrea A. Chiba, and Michela Gallagher. Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. *Nature Neuroscience*, 1(2):155–159, June 1998. Publisher: Nature Publishing Group.

[82] Jung Hoon Sul, Hoseok Kim, Namjung Huh, Daeyeol Lee, and Min Whan Jung. Distinct Roles of Rodent Orbitofrontal and Medial Prefrontal Cortex in Decision Making. *Neuron*, 66(3):449–460, May 2010. Publisher: Elsevier.

[83] Martin O'Neill and Wolfram Schultz. Coding of Reward Risk by Orbitofrontal Neurons Is Mostly Distinct from Coding of Reward Value. *Neuron*, 68(4):789–800, November 2010. Publisher: Elsevier.

[84] C. Padoa-Schioppa. Neurobiology of Economic Choice: A Good-Based Model. *Annual Review of Neuroscience*, 34:333–359, 2011.

[85] Steven W. Kennerley, Aspandiar F. Dahmubed, Antonio H. Lara, and Jonathan D. Wallis. Neurons in the Frontal Lobe Encode the Value of Multiple Decision Variables. *Journal of cognitive neuroscience*, 21(6):1162–1178, June 2009.

[86] C. Amiez, J.P. Joseph, and E. Procyk. Reward Encoding in the Monkey Anterior Cingulate Cortex. *Cerebral Cortex*, 16(7):1040–1055, July 2006.

[87] Matthew F. S. Rushworth and Timothy E. J. Behrens. Choice, uncertainty and value in prefrontal and cingulate cortex. *Nature Neuroscience*, 11(4):389–397, April 2008. Publisher: Nature Publishing Group.

[88] Munetaka Shidara and Barry J. Richmond. Anterior Cingulate: Single Neuronal Signals Related to Degree of Reward Expectancy. *Science*, 296(5573):1709–1711, May 2002. Publisher: American Association for the Advancement of Science.

[89] Steven W. Kennerley, Timothy E. J. Behrens, and Jonathan D. Wallis. Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nature neuroscience*, 14(12):1581–1589, October 2011.

[90] Xinying Cai and Camillo Padoa-Schioppa. Neuronal Encoding of Subjective Value in Dorsal and Ventral Anterior Cingulate Cortex. *Journal of Neuroscience*, 32(11):3791–3808, March 2012. Publisher: Society for Neuroscience Section: Articles.

[91] Nils Kolling, Marco K. Wittmann, Tim E. J. Behrens, Erie D. Boorman, Rogier B. Mars, and Matthew F. S. Rushworth. Value, search, persistence and model updating in anterior cingulate cortex. *Nature Neuroscience*, 19(10):1280–1285, October 2016. Publisher: Nature Publishing Group.

[92] Thomas Akam, Ines Rodrigues-Vaz, Ivo Marcelo, Xiangyu Zhang, Michael Pereira, Rodrigo Freire Oliveira, Peter Dayan, and Rui M. Costa. The Anterior Cingulate Cortex Predicts Future States to Mediate Model-Based Action Selection. *Neuron*, 109(1):149–163.e7, January 2021.

[93] R. Becket Ebitz and Benjamin Yost Hayden. Dorsal anterior cingulate: a Rorschach test for cognitive neuroscience. *Nature Neuroscience*, 19(10):1278–1279, October 2016. Publisher: Nature Publishing Group.

[94] Armin Lak, Michael Okun, Morgane M. Moss, Harsha Gurnani, Karolina Farrell, Miles J. Wells, Charu Bai Reddy, Adam Kepecs, Kenneth D. Harris, and Matteo Carandini. Dopaminergic and Prefrontal Basis of Learning from Sensory Confidence and Reward Value. *Neuron*, 0(0), December 2019.

[95] Laura H Corbit and Bernard W Balleine. The role of prelimbic cortex in instrumental conditioning. *Behavioural Brain Research*, 146(1):145–157, November 2003.

[96] Jennifer R. St. Onge and Stan B. Floresco. Prefrontal Cortical Contribution to Risk-Based Decision Making. *Cerebral Cortex*, 20(8):1816–1828, August 2010.

[97] Antonio Rangel, Colin Camerer, and P. Read Montague. A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, 9(7):545–556, July 2008. Publisher: Nature Publishing Group.

[98] Tomoya Ohnuki, Yuma Osako, Hiroyuki Manabe, Yoshio Sakurai, and Junya Hirokawa. Over-representation of fundamental decision variables in the prefrontal cortex underlies decision bias. *Neuroscience Research*, 173:1–13, December 2021.

[99] Jonathan D Wallis and Steven W Kennerley. Heterogeneous reward signals in prefrontal cortex. *Current Opinion in Neurobiology*, 20(2):191–198, April 2010.

[100] Steven W. Kennerley and Jonathan D. Wallis. Evaluating choices by single neurons in the frontal lobe: outcome value encoded across multiple decision variables. *European Journal of Neuroscience*, 29(10):2061–2073, 2009. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1460-9568.2009.06743.x.

[101] Steven W. Kennerley and Mark E. Walton. Decision Making and Reward in Frontal Cortex. *Behavioral Neuroscience*, 125(3):297–317, June 2011.

[102] Michael L. Platt and Paul W. Glimcher. Neural correlates of decision variables in parietal cortex. *Nature*, 400(6741):233–238, July 1999. Publisher: Nature Publishing Group.

[103] Takayuki Hosokawa, Steven W. Kennerley, Jennifer Sloan, and Jonathan D. Wallis. Single-Neuron Mechanisms Underlying Cost-Benefit Analysis in Frontal Cortex. *Journal of Neuroscience*, 33(44):17385–17397, October 2013. Publisher: Society for Neuroscience Section: Articles.

[104] David J.-N. Maisson, Tyler V. Cash-Padgett, Maya Z. Wang, Benjamin Y. Hayden, Sarah R. Heilbronner, and Jan Zimmermann. Choice-relevant information transformation along a ventrodorsal axis in the medial prefrontal cortex. *Nature Communications*, 12(1):4830, August 2021. Publisher: Nature Publishing Group.

[105] James M. Otis, Vijay M. K. Namboodiri, Ana M. Matan, Elisa S. Voets, Emily P. Mohorn, Oksana Kosyk, Jenna A. McHenry, J. Elliott Robinson, Shanna L. Resendez, Mark A. Rossi, and Garret D. Stuber. Prefrontal cortex output circuits guide reward seeking through divergent cue encoding. *Nature*, 543(7643):103–107, March 2017. Number: 7643 Publisher: Nature Publishing Group.

[106] Gregor Rainer, Wael F. Asaad, and Earl K. Miller. Selective representation of relevant information by neurons in the primate prefrontal cortex. *Nature*, 393(6685):577–579, June 1998. Publisher: Nature Publishing Group.

[107] Liqun Luo, Edward M. Callaway, and Karel Svoboda. Genetic Dissection of Neural Circuits: A Decade of Progress. *Neuron*, 98(2):256–281, April 2018.

[108] Ofer Yizhar, Lief E. Fenno, Thomas J. Davidson, Murtaza Mogri, and Karl Deisseroth. Optogenetics in Neural Systems. *Neuron*, 71(1):9–34, July 2011. Publisher: Elsevier.

[109] Parisa Mahmoudi, Hadi Veladi, and Firooz G. Pakdel. Optogenetics, Tools and Applications in Neurobiology. *Journal of Medical Signals and Sensors*, 7(2):71–79, 2017.

[110] Paula L. Croxson, Heidi Johansen-Berg, Timothy E. J. Behrens, Matthew D. Robson, Mark A. Pinsk, Charles G. Gross, Wolfgang Richter, Marlene C. Richter, Sabine Kastner, and Matthew F. S. Rushworth. Quantitative Investigation of Connections of the Prefrontal Cortex in the Human and Macaque using Probabilistic Diffusion Tractography. *Journal of Neuroscience*, 25(39):8854–8866, September 2005. Publisher: Society for Neuroscience Section: Behavioral/Systems/Cognitive.

[111] D. Kvitsiani, S. Ranade, B. Hangya, H. Taniguchi, J. Z. Huang, and A. Kepecs. Distinct behavioural and network correlates of two interneuron types in prefrontal cortex. *Nature*, 498(7454):363–366, June 2013. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 7454 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Decision Subject_term_id: decision.

[112] Lucas Pinto and Yang Dan. Cell-Type-Specific Activity in Prefrontal Cortex during Goal-Directed Behavior. *Neuron*, 87(2):437–450, July 2015.

[113] S. Ciocchi, J. Passecker, H. Malagon-Vina, N. Mikus, and T. Klausberger. Selective information routing by ventral hippocampal CA1 projection neurons. *Science*, 348(6234):560–563, May 2015.

[114] Takuma Kitanishi, Ryoko Umaba, and Kenji Mizuseki. Robust information routing by dorsal subiculum neurons. *Science Advances*, 7(11):eabf1913, March 2021.

[115] Junya Hirokawa, Alexander Vaughan, Paul Masset, Torben Ott, and Adam Kepecs. Frontal cortex neuron types categorically encode single decision variables. *Nature*, 576(7787):446–451, December 2019. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 7787 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Decision;Neural encoding Subject_term_id: decision;neural-encoding.

[116] Eric Burguière, Patrícia Monteiro, Guoping Feng, and Ann M. Graybiel. Optogenetic Stimulation of Lateral Orbitofronto-Striatal Pathway Suppresses Compulsive Behaviors. *Science (New York, N.Y.)*, 340(6137):10.1126/science.1232380, June 2013.

[117] Brian Zingg, Hong-Wei Dong, Huizhong Whit Tao, and Li I. Zhang. Input–output organization of the mouse claustrum. *Journal of Comparative Neurology*, 526(15):2428–2443, 2018. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cne.24502.

[118] Paul G Anastasiades, Christina Boada, and Adam G Carter. Cell-Type-Specific D1 Dopamine Receptor Modulation of Projection Neurons and Interneurons in the Prefrontal Cortex. *Cerebral Cortex*, 29(7):3224–3242, July 2019.

[119] Henry H. Yin, Sean B. Ostlund, Barbara J. Knowlton, and Bernard W. Balleine. The role of the dorsomedial striatum in instrumental conditioning. *European Journal of Neuroscience*, 22(2):513–523, 2005. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1460-9568.2005.04218.x.

[120] Bernard W. Balleine, Mauricio R. Delgado, and Okihide Hikosaka. The Role of the Dorsal Striatum in Reward and Decision-Making. *Journal of Neuroscience*, 27(31):8161–8165, August 2007. Publisher: Society for Neuroscience Section: Mini-Review.

[121] Bernard W Balleine and John P O'Doherty. Human and Rodent Homologies in Action Control: Corticostriatal Determinants of Goal-Directed and Habitual Action. *Neuropsychopharmacology*, 35(1):48–69, January 2010.

[122] Jackson D. Schumacher, Mieke van Holstein, Vaishali Bagrodia, Hannah B. Le Bouder, and Stan B. Floresco. Dorsomedial striatal contributions to different forms of risk/reward decision making. *Neurobiology of Learning and Memory*, 178:107369, February 2021.

[123] Thomas A. Stalnaker, Gwen G. Calhoon, Masaaki Ogawa, Matthew R. Roesch, and Geoffrey Schoenbaum. Neural correlates of stimulus-response and response-outcome associations in dorsolateral versus dorsomedial striatum. *Frontiers in Integrative Neuroscience*, 4, May 2010. Publisher: Frontiers.

[124] Henry H. Yin, Barbara J. Knowlton, and Bernard W. Balleine. Blockade of NMDA receptors in the dorsomedial striatum prevents action–outcome learning in instrumental conditioning. *European Journal of Neuroscience*, 22(2):505–512, 2005. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1460-9568.2005.04219.x.

[125] Henry H. Yin, Sean B. Ostlund, and Bernard W. Balleine. Reward-guided learning beyond dopamine in the nucleus accumbens: The integrative functions of cortico-basal ganglia networks. *The European journal of neuroscience*, 28(8):1437–1448, October 2008.

[126] John P. O'Doherty, Ralf Deichmann, Hugo D. Critchley, and Raymond J. Dolan. Neural Responses during Anticipation of a Primary Taste Reward. *Neuron*, 33(5):815–826, February 2002. Publisher: Elsevier.

[127] Esther Aarts, Deanna L. Wallace, Linh C. Dang, William J. Jagust, Roshan Cools, and Mark D'Esposito. Dopamine and the Cognitive Downside of a Promised Bonus. *Psychological Science*, 25(4):1003–1009, April 2014. Publisher: SAGE Publications Inc.

[128] Oliver J. Robinson, Michael J. Frank, Barbara J. Sahakian, and Roshan Cools. Dissociable responses to punishment in distinct striatal regions during reversal learning. *Neuroimage*, 51(4-4):1459–1467, July 2010.

[129] Xinying Cai, Soyoun Kim, and Daeyeol Lee. Heterogeneous Coding of Temporally Discounted Values in the Dorsal and Ventral Striatum during Intertemporal Choice. *Neuron*, 69(1):170–182, January 2011. Publisher: Elsevier.

[130] Carinna M. Torgerson, Andrei Irimia, S. Y. Matthew Goh, and John Darrell Van Horn. The DTI connectivity of the human claustrum. *Human Brain Mapping*, 36(3):827–838, October 2014.

[131] Ryan Remedios, Nikos K. Logothetis, and Christoph Kayser. Unimodal Responses Prevail within the Multisensory Claustrum. *Journal of Neuroscience*, 30(39):12902–12907, September 2010. Publisher: Society for Neuroscience Section: Brief Communications.

[132] Ryan Remedios, Nikos K. Logothetis, and Christoph Kayser. A role of the claustrum in auditory scene analysis by reflecting sensory change. *Frontiers in Systems Neuroscience*, 0, 2014. Publisher: Frontiers.

[133] Francis C Crick and Christof Koch. What is the function of the claustrum? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1458):1271–1279, June 2005.

[134] Gal Atlan, Anna Terem, Noa Peretz-Rivlin, Maya Groysman, and Ami Citri. Mapping synaptic cortico-claustral connectivity in the mouse. *Journal of Comparative Neurology*, 525(6):1381–1402, 2017. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cne.23997.

[135] Jesse Jackson, Mahesh M. Karnani, Boris V. Zemelman, Denis Burdakov, and Albert K. Lee. Inhibitory Control of Prefrontal Cortex by the Claustrum. *Neuron*, 99(5):1029–1039.e4, September 2018.

[136] Ethan G. McBride, Saurabh R. Gandhi, Jacqulyn R. Kuyat, Douglas R. Ollerenshaw, Anton Arkhipov, Christof Koch, and Shawn R. Olsen. Influence of claustrum on cortex varies by area, layer, and cell type. *Neuron*, 111(2):275–290.e5, January 2023. Publisher: Elsevier.

[137] Yael Goll, Gal Atlan, and Ami Citri. Attention: the claustrum. *Trends in Neurosciences*, 38(8):486–495, August 2015.

[138] Brian N. Mathur. The claustrum in review. *Frontiers in Systems Neuroscience*, 8, April 2014.

[139] Jianfeng Liu, Ruyan Wu, Bernard Johnson, Jimmy Vu, Caroline Bass, and Jun-Xu Li. The Claustrum-Prefrontal Cortex Pathway Regulates Impulsive-Like Behavior. *Journal of Neuroscience*, 39(50):10071–10080, December 2019. Publisher: Society for Neuroscience Section: Research Articles.

[140] Gal Atlan, Anna Terem, Noa Peretz-Rivlin, Kamini Sehrawat, Ben Jerry Gonzales, Guy Pozner, Gen-ichi Tasaka, Yael Goll, Ron Refaeli, Ori Zviran, Byung Kook Lim, Maya Groysman, Inbal Goshen, Adi Mizrahi, Israel Nelken, and Ami Citri. The Claustrum Supports Resilience to Distraction. *Current Biology*, 28(17):2752–2762.e7, September 2018. Publisher: Elsevier.

[141] Gal Atlan, Noa Matosevich, Noa Peretz-Rivlin, Idit Marsh-Yvgi, Noam Zelinger, Eden Chen, Timna Kleinman, Noa Bleistein, Efrat Sheinbach, Maya Groysman, Yuval Nir, and Ami Citri. Claustrum neurons projecting to the anterior cingulate restrict engagement during sleep and behavior. *Nature Communications*, 15(1):5415, June 2024. Publisher: Nature Publishing Group.

[142] Jared B. Smith, Glenn D. R. Watson, Zhifeng Liang, Yikang Liu, Nanyin Zhang, and Kevin D. Alloway. A Role for the Claustrum in Salience Processing? *Frontiers in Neuroanatomy*, 13, 2019. Publisher: Frontiers.

[143] Anna Terem, Ben Jerry Gonzales, Noa Peretz-Rivlin, Reut Ashwal-Fluss, Noa Bleistein, Maria del Mar Reus-Garcia, Diptendu Mukherjee, Maya Groysman, and Ami Citri. Claustral Neurons Projecting to Frontal Cortex Mediate Contextual Association of Reward. *Current Biology*, 30(18):3522–3532.e6, September 2020.

[144] Michael G. White, Matthew Panicker, Chaoqi Mu, Ashley M. Carter, Bradley M. Roberts, Poorna A. Dharmasri, and Brian N. Mathur. Anterior Cingulate Cortex Input to the Claustrum Is Required for Top-Down Action Control. *Cell Reports*, 22(1):84–95, January 2018. Publisher: Elsevier.

[145] Leon Fodoulian, Olivier Gschwend, Chieko Huber, Sophie Mutel, Rodrigo F. Salazar, Roberta Leone, Jean-Rodolphe Renfer, Kazadi Ekundayo, Ivan Rodriguez, and Alan Carleton. The claustrum-medial prefrontal cortex network controls attentional set-shifting. *bioRxiv*, page 2020.10.14.339259, October 2020. Publisher: Cold Spring Harbor Laboratory Section: New Results.

[146] Michael G. White, Chaoqi Mu, Houman Qadir, Maxwell Madden, Hongkui Zeng, and Brian N. Mathur. The mouse claustrum is required for optimal behavioral performance under high cognitive demand. *Biological psychiatry*, 88(9):719–726, November 2020.

[147] Solange P. Brown, Brian N. Mathur, Shawn R. Olsen, Pierre-Hervé Luppi, Martha E. Bickford, and Ami Citri. New Breakthroughs in Understanding the Role of Functional Interactions between the Neocortex and the Claustrum. *The Journal of Neuroscience*, 37(45):10877–10881, November 2017.

[148] Nathan C Klapoetke, Yasunobu Murata, Sung Soo Kim, Stefan R. Pulver, Amanda Birdsey-Benson, Yong Ku Cho, Tania K Morimoto, Amy S Chuong, Eric J Carpenter, Zhijian Tian, Jun Wang, Yinlong Xie, Zhixiang Yan, Yong Zhang, Brian Y Chow, Barbara Surek, Michael Melkonian, Vivek Jayaraman, Martha Constantine-Paton, Gane Ka-Shu Wong, and Edward S Boyden. Independent Optical Excitation of Distinct Neural Populations. *Nature methods*, 11(3):338–346, March 2014.

[149] Christopher P. Burgess, Armin Lak, Nicholas A. Steinmetz, Peter Zatka-Haas, Charu Bai Reddy, Elina A.K. Jacobs, Jennifer F. Linden, Joseph J. Paton, Adam Ranson, Sylvia Schröder, Sofia Soares, Miles J. Wells, Lauren E. Wool, Kenneth D. Harris, and Matteo Carandini. High-Yield Methods for Accurate Two-Alternative Visual Psychophysics in Head-Fixed Mice. *Cell Reports*, 20(10):2513–2524, September 2017.

[150] James J. Jun, Nicholas A. Steinmetz, Joshua H. Siegle, Daniel J. Denman, Marius Bauza, Brian Barbarits, Albert K. Lee, Costas A. Anastassiou, Alexandru Andrei, Çağatay Aydın, Mladen Barbic, Timothy J. Blanche, Vincent Bonin, João Couto, Barundeb Dutta, Sergey L. Gratiy, Diego A. Gutnisky, Michael Häusser, Bill Karsh, Peter Ledochowitsch, Carolina Mora Lopez, Catalin Mitelut, Silke Musa, Michael Okun, Marius Pachitariu, Jan Putzeys, P. Dylan Rich, Cyrille Rossant, Wei-lung Sun, Karel Svoboda, Matteo Carandini, Kenneth D. Harris, Christof Koch, John O'Keefe,

and Timothy D. Harris. Fully integrated silicon probes for high-density recording of neural activity. *Nature*, 551(7679):232–236, November 2017.

[151] Marius Pachitariu, Nicholas Steinmetz, Shabnam Kadir, Matteo Carandini, and Harris Kenneth D. Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. preprint, Neuroscience, June 2016.

[152] Quanxin Wang, Song-Lin Ding, Yang Li, Josh Royall, David Feng, Phil Lesnar, Nile Graddis, Maitham Naeemi, Benjamin Facer, Anh Ho, Tim Dolbeare, Brandon Blanchard, Nick Dee, Wayne Wakeman, Karla E. Hirokawa, Aaron Szafer, Susan M. Sunkin, Seung Wook Oh, Amy Bernard, John W. Phillips, Michael Hawrylycz, Christof Koch, Hongkui Zeng, Julie A. Harris, and Lydia Ng. The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas. *Cell*, 181(4):936–953.e20, May 2020.

[153] Kristoffer Lauridsen, Annie Ly, Emily D. Prévost, Connor McNulty, Dillon J. McGovern, Jian Wei Tay, Joseph Dragavon, and David H. Root. A Semi-Automated Workflow for Brain Slice Histology Alignment, Registration, and Cell Quantification (SHARCQ). *eNeuro*, 9(2), March 2022. Publisher: Society for Neuroscience Section: Open Source Tools and Methods.

[154] Il Memming Park, Miriam L. R. Meister, Alexander C. Huk, and Jonathan W. Pillow. Encoding and decoding in parietal cortex during sensorimotor decision-making. *Nature Neuroscience*, 17(10):1395–1403, October 2014. Publisher: Nature Publishing Group.

[155] N. A. Steinmetz, P. Zatka-Haas, M. Carandini, and K. D. Harris. Distributed coding of choice, action and engagement across the mouse brain. *Nature*, 576(7786):266–273, December 2019. Edition: 2019/11/30.

[156] J-R. King and S. Dehaene. Characterizing the dynamics of mental representations: the temporal generalization method. *Trends in cognitive sciences*, 18(4):203–210, April 2014.

[157] Eric Maris and Robert Oostenveld. Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1):177–190, August 2007.

[158] Stefan Haufe, Frank Meinecke, Kai Görgen, Sven Dähne, John-Dylan Haynes, Benjamin Blankertz, and Felix Bießmann. On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, 87:96–110, February 2014.

[159] Daniel Bernoulli. Exposition of a New Theory on the Measurement of Risk. *Econometrica*, 22(1):23–36, 1954. Publisher: [Wiley, Econometric Society].

[160] Amos Tversky. A Critique of Expected Utility Theory: Descriptive and Normative Considerations. *Erkenntnis (1975-)*, 9(2):163–173, 1975. Publisher: Springer.

[161] Nicholas C Barberis. Thirty Years of Prospect Theory in Economics: A Review and Assessment. *Journal of Economic Perspectives*, 27(1):173–196, February 2013.

[162] Matteo Carandini and Anne K. Churchland. Probing perceptual decisions in rodents. *Nature Neuroscience*, 16(7):824–831, July 2013. Publisher: Nature Publishing Group.

[163] Chaofei Bao, Xiaoyue Zhu, Joshua Mōller-Mara, Jingjie Li, Sylvain Dubroqua, and Jeffrey C. Erlich. The rat frontal orienting field dynamically encodes value for economic decisions under risk. *Nature Neuroscience*, 26(11):1942–1952, November 2023. Publisher: Nature Publishing Group.

[164] Christine M. Constantinople, Alex T. Piet, and Carlos D. Brody. An Analysis of Decision under Risk in Rats. *Current Biology*, 29(12):2066–2074.e5, June 2019.

[165] Masaru Kuwabara, Ningdong Kang, Timothy E Holy, and Camillo Padoa-Schioppa. Neural mechanisms of economic choices in mice. *eLife*, 9:e49669, February 2020. Publisher: eLife Sciences Publications, Ltd.

[166] Matthew R. Roesch, Donna J. Calu, and Geoffrey Schoenbaum. Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nature Neuroscience*, 10(12):1615–1624, December 2007. Publisher: Nature Publishing Group.

[167] Reza Shadmehr, Thomas R. Reppert, Erik M. Summerside, Tehrim Yoon, and Alaa A. Ahmed. Movement vigor as a reflection of subjective economic utility. *Trends in neurosciences*, 42(5):323–336, May 2019.

[168] Samuel Liebana Garcia, Aeron Laffere, Chiara Toschi, Louisa Schilling, Jacek Podlaski, Matthias Fritsche, Peter Zatka-Haas, Yulong Li, Rafal Bogacz, Andrew Saxe, and Armin Lak. Striatal dopamine reflects individual long-term learning trajectories, December 2023. Pages: 2023.12.14.571653 Section: New Results.

[169] William R. Stauffer, Armin Lak, Peter Bossaerts, and Wolfram Schultz. Economic Choices Reveal Probability Distortion in Macaque Monkeys. *Journal of Neuroscience*, 35(7):3146–3154, February 2015. Publisher: Society for Neuroscience Section: Articles.

[170] Mark H. Histed, Lauren A. Carvalho, and John H. R. Maunsell. Psychophysical measurement of contrast sensitivity in the behaving mouse. *Journal of Neurophysiology*, 107(3):758–765, February 2012. Publisher: American Physiological Society.

[171] Armin Lak, Gil M. Costa, Erin Romberg, Alexei A. Koulakov, Zachary F. Mainen, and Adam Kepecs. Orbitofrontal Cortex Is Required for Optimal Waiting Based on Decision Confidence. *Neuron*, 84(1):190–201, October 2014.

[172] Sandra Reinert, Mark Hübener, Tobias Bonhoeffer, and Pieter M. Goltstein. Mouse prefrontal cortex represents learned rules for categorization. *Nature*, 593(7859):411–417, May 2021. Publisher: Nature Publishing Group.

[173] Avinash R. Vaidya and Lesley K. Fellows. Testing necessary regional frontal contributions to value assessment and fixation-based updating. *Nature Communications*, 6(1):10120, December 2015. Number: 1 Publisher: Nature Publishing Group.

[174] Dino J. Levy and Paul W. Glimcher. Comparing Apples and Oranges: Using Reward-Specific and Reward-General Subjective Value Representation in the Brain. *Journal of Neuroscience*, 31(41):14693–14707, October 2011. Publisher: Society for Neuroscience Section: Articles.

[175] Timothy E. J. Behrens, Mark W. Woolrich, Mark E. Walton, and Matthew F. S. Rushworth. Learning the value of information in an uncertain world. *Nature Neuroscience*, 10(9):1214–1221, September 2007. Publisher: Nature Publishing Group.

[176] Matthew R. Roesch and Carl R. Olson. Neuronal Activity in Primate Orbitofrontal Cortex Reflects the Value of Time. *Journal of Neurophysiology*, 94(4):2457–2471, October 2005. Publisher: American Physiological Society.

[177] William E. Allen, Michael Z. Chen, Nandini Pichamoorthy, Rebecca H. Tien, Marius Pachitariu, Liqun Luo, and Karl Deisseroth. Thirst regulates motivated behavior through modulation of brainwide neural population dynamics. *Science*, 364(6437):eaav3932, April 2019. Publisher: American Association for the Advancement of Science.

[178] A. Lak, E. Hueske, J. Hirokawa, P. Masset, T. Ott, A. E. Urai, T. H. Donner, M. Carandini, S. Tonegawa, N. Uchida, and A. Kepecs. Reinforcement biases subsequent perceptual decisions when confidence is low, a widespread behavioral phenomenon. *Elife*, 9, April 2020. Edition: 2020/04/15.

[179] Peter Y. Wang, Cristian Boboila, Matthew Chin, Alexandra Higashi-Howard, Philip Shamash, Zheng Wu, Nicole P. Stein, L. F. Abbott, and Richard Axel. Transient and Persistent Representations of Odor Value in Prefrontal Cortex. *Neuron*, 108(1):209–224.e6, October 2020. Publisher: Elsevier.

[180] Vyacheslav R. Karolis, Maurizio Corbetta, and Michel Thiebaut de Schotten. The architecture of functional lateralisation and its relationship to callosal connectivity in the human brain. *Nature Communications*, 10(1):1417, March 2019. Publisher: Nature Publishing Group.

[181] Edward H. Silson, Peter Zeidman, Tomas Knapen, and Chris I. Baker. Representation of Contralateral Visual Space in the Human Hippocampus. *The Journal of Neuroscience*, 41(11):2382–2392, March 2021.

[182] Sarah R. Heilbronner and Benjamin Y. Hayden. Dorsal Anterior Cingulate Cortex: A Bottom-Up View. *Annual Review of Neuroscience*, 39(1):149–170, July 2016.

[183] Nikolaus Kriegeskorte and Pamela K Douglas. Interpreting encoding and decoding models. *Current Opinion in Neurobiology*, 55:167–179, April 2019.

[184] Michael E Ragozzino, Jane Jih, and Arianna Tzavos. Involvement of the dorsomedial striatum in behavioral flexibility: role of muscarinic cholinergic receptors. *Brain Research*, 953(1):205–214, October 2002.

[185] Bernard W. Balleine. Neural bases of food-seeking: Affect, arousal and reward in corticostriatolimbic circuits. *Physiology & Behavior*, 86(5):717–730, December 2005.

[186] Maxime Chevée, Eric A. Finkel, Su-Jeong Kim, Daniel H. O'Connor, and Solange P. Brown. Neural activity in the mouse claustrum in a cross-modal sensory selection task. *Neuron*, 110(3):486–501.e7, February 2022. Publisher: Elsevier.

[187] Dechen Liu, Juan Deng, Zhewei Zhang, Zhi-Yu Zhang, Yan-Gang Sun, Tianming Yang, and Haishan Yao. Orbitofrontal control of visual cortex gain promotes visual associative learning. *Nature Communications*, 11(1):2784, June 2020. Publisher: Nature Publishing Group.

[188] Barbara J Hunnicutt, Bart C Jongbloets, William T Birdsong, Katrina J Gertz, Haining Zhong, and Tianyi Mao. A comprehensive excitatory input map of the striatum reveals novel functional organization. *eLife*, 5:e19103, November 2016. Publisher: eLife Sciences Publications, Ltd.

[189] Kyuhyun Choi, Eugenio Piasini, Edgar Díaz-Hernández, Luigim Vargas Cifuentes, Nathan T. Henderson, Elizabeth N. Holly, Manivannan Subramaniyan, Charles R. Gerfen, and Marc V. Fuccillo. Distributed processing for value-based choice by prelimbic circuits targeting anterior-posterior dorsal striatal subregions in male mice. *Nature Communications*, 14(1):1920, April 2023. Publisher: Nature Publishing Group.

[190] Benjamin Y. Hayden and Michael L. Platt. Neurons in Anterior Cingulate Cortex Multiplex Information about Reward and Action. *Journal of Neuroscience*, 30(9):3339–3346, March 2010. Publisher: Society for Neuroscience Section: Articles.

[191] Chung-Hay Luk and Jonathan D. Wallis. Choice Coding in Frontal Cortex during Stimulus-Guided or Action-Guided Decision-Making. *Journal of Neuroscience*, 33(5):1864–1871, January 2013. Publisher: Society for Neuroscience Section: Articles.

[192] Amitai Shenhav, Matthew M. Botvinick, and Jonathan D. Cohen. The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, 79(2):217–240, July 2013.

[193] Shih-Yi Tseng, Selmaan N. Chettih, Charlotte Arlt, Roberto Barroso-Luque, and Christopher D. Harvey. Shared and specialized coding across posterior cortical areas for dynamic navigation decisions. *Neuron*, 110(15):2484–2502.e16, August 2022. Publisher: Elsevier.

[194] Kerstin Preuschoff, Peter Bossaerts, and Steven R. Quartz. Neural Differentiation of Expected Reward and Risk in Human Subcortical Structures. *Neuron*, 51(3):381–390, August 2006. Publisher: Elsevier.

[195] Marci R Mitchell, Virginia G Weiss, B Sofia Beas, Drake Morgan, Jennifer L Bizon, and Barry Setlow. Adolescent Risk Taking, Cocaine Self-Administration, and Striatal Dopamine Signaling. *Neuropsychopharmacology*, 39(4):955–962, March 2014.

[196] Mattia Rigotti, Omri Barak, Melissa R. Warden, Xiao-Jing Wang, Nathaniel D. Daw, Earl K. Miller, and Stefano Fusi. The importance of mixed selectivity in complex cognitive tasks. *Nature*, 497(7451):585–590, May 2013. Publisher: Nature Publishing Group.

[197] Susana Q. Lima, Tomáš Hromádka, Petr Znamenskiy, and Anthony M. Zador. PINP: A New Method of Tagging Neuronal Populations for Identification during In Vivo Electrophysiological Recording. *PLOS ONE*, 4(7):e6099, July 2009. Publisher: Public Library of Science.

[198] Karl Deisseroth, Guoping Feng, Ania K. Majewska, Gero Miesenböck, Alice Ting, and Mark J. Schnitzer. Next-Generation Optical Technologies for Illuminating Genetically Targeted Brain Circuits. *Journal of Neuroscience*, 26(41):10380–10386, October 2006. Publisher: Society for Neuroscience Section: Symposia and Mini-Symposia.

[199] Ichiji Tasaki. Collision of two nerve impulses in the nerve fibre. *Biochimica et Biophysica Acta*, 3:494–497, January 1949.

[200] Keita Mitani, Masanori Kawabata, Yoshikazu Isomura, and Yutaka Sakai. Automated and parallelized spike collision tests to identify spike signal projections. *iScience*, 25(10), October 2022. Publisher: Elsevier.

[201] Hyeonhui Kim, Minki Kim, Sun-Kyoung Im, and Sungsoon Fang. Mouse Cre-LoxP system: general principles to determine tissue-specific roles of target genes. *Laboratory Animal Research*, 34(4):147–159, December 2018.